

STIC-Biotech/ChemLib

From: Duffy, Patricia
Sent: Tuesday, June 19, 2001 12:23 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search 09/438,185
Importance: High

In re: 09/438,185

Please search SEQ ID NO:1047.
Please backtranslate SEQ ID NO:1047 into NA and run against NA databases.
Please perform an interference search and print out top 25 hits.

Thank you,
Patricia A. Duffy
Art Unit 1645
Office CM-1, 8D05
Mailbox CM-1, 8E12
Phone 703-305-7555

Point of Contact:
Mona Smith
Technical Info. Specialist
CM1 12C14 Tel: 308-3278

Point of Contact:
Mona Smith
Technical Info. Specialist
CM1 12C14 Tel: 308-3278

XX	AAV34584-Y35879 represent the proteins encoded by all the open reading	
CCC	frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.	
CCC	C. pneumoniae causes respiratory disease such as pneumonia and	
CCC	bronchitis and is thought to be a contributing factor in heart	
CCC	disease, sarcoidosis, sinusitis, purulent otitis media, erythema	
CCC	nodosum or pharyngitis. The polypeptides encoded by the open reading	
CCC	frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in	
CCC	immunogenic compositions as vaccines. Vectors containing C. pneumoniae	
CCC	nucleotide sequences can also be used as immunogenic compositions,	
CCC	especially where the vector directs the expression of a neutralising	
CCC	epitope of C. pneumoniae.	
XX	Sequence	259 AA:
XX		

Query Match	66.24;	Score 1250.5;	DB 20;	Length 259;
at Local Similarity	91.8%;	Pred. NO. 3.7e-118;		
Ches 245;	Conservative	3;	Mismatches 8;	Indels 11; Gaps 1;
Qy	96	VVSTPPNNMLVYRLLSSRFSLWKSYPCEFFLDYLEAFGLSDFDLHQAVIKFELETHF	155	
Db	4	mvsctfvtfsmekl-----kfldyleafglsdfldhqavikfelethf	52	
Qy	156	SYYPVSGFVAPHQYLSLQDRYFPIASVMRTLKDKNFSLTPDLIHDLHGVPMLLPSPS	215	
Db	53	syypvsgfvaphqyislqdrfyfiavsmrtldkdnfsitpdlidhllgvpwllhpsts	112	
Qy	216	EFFINMGRLFTVIEKVOALPSKKORTOTLOSNLIAIVRCFWFTVESGLIENHGKRAYG	275	
Db	113	effinmgrftvkiekvalpskkgrigtqlsqnlaiavrcftwftvesglienhegrkayg	172	
Qy	276	AVLISSPQELGHAFIDNVRVLPLELDGIIRLPNTSTPOETLSIRIHFDELVELTSKLEW	335	
Db	173	avlisspqeighafidnvrvipleldgiiirpntstpqetlfsirhfdelveltsklew	232	
Qy	336	MLDOGLLESIPYNOEKVLSGFVEVLQ	362	
Db	233	mldggllesipynqekvlsqfvevlcq	259	

RESULT	2	
XX	AAR50185	
XX	ID AAR50185 standard; Protein; 498 AA.	
XX		
XX	AAR50185;	
XX		
XX	04-NOV-1994 (first entry)	
XX		
XX	Rat tyrosine hydroxylase.	
XX		
XX	Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;	
XX	enhanced enzymatic activity; wild-type; transformation; dopamine;	
XX	Parkinsons disease; Alzheimers disease; brain; encapsulation;	
XX	selectively permeable; polymer capsule; antibody.	
XX		
OS	Rattus rattus.	
XX		
XX	US5300436-A.	
XX		
XX	05-APR-1994.	
XX		
XX	13-MAR-1991; 91US-0669446.	
XX		
XX	13-MAR-1991; 91US-0669446.	
PR	26-JAN-1993; 93US-0009075.	
XX		
XX	(UYNV) UNIV NEW YORK STATE.	
XX		
XX	Filler D, Friedhoff AJ, Goldstein M, Wu J;	
PI		
PI	WPI; 1994-125849/15.	
XX		
XX		

Modified DNA encoding variant tyrosine hydroxylase with an N-terminal amino acid subseq. and cells contg. it - can be used to treat diseases associated with defective function of the enzyme, e.g. Parkinson's disease or Alzheimer's disease

Disclosure: Flig 5; 27pp; English.

This sequence represents rat tyrosine hydroxylase (TH) and may be used in the production of the variant TH molecules of the invention. These variants contain amino acid substitutions in the N-terminal portion of the protein and the Ser31 molecule these substitutions are at positions Ser8, Ser19, Ser31, Arg38, Ser40, Glu43 or Arg46. Variants containing one or more of these amino acid substitutions have substantially enhanced enzymatic activity compared to the wild-type enzyme. The most pref. substitution being Ser40 for Tyr or Leu. DNA encoding the TH variants, and cells transformed with this DNA may be used for treating Parkinson's disease with defective function of TH, or dopamine, e.g. Parkinson's disease and Alzheimers disease, and effective disorders. The cells can then be implanted into the brain or encapsulated in a selectively permeable polymer capsule which allows release of the cells products but protects them from attack by the hosts antibodies or cells.

Sequence 498 AA.

```

Query Match      11.88;  Score 222;  DB 15;  Length 498;
Best Local Similarity 26.7%;  Pred. No. 7.1e-14;
Matches 65;  Conservative 39;  Mismatches 103;  Indels 36;  Gaps

QY 112 SSRESLWKS-----YCPRFIDYLAEAGLLSDPLDH-----QAVIKFELETHF 155
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 227 aeeiatwkevvtlkglyathacrehlegqllerycyredsipledvsrflkertg 286
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 156 SYYPVSGFVAPHQYLSLQDRFFPIASVMETLKDKNFSLTPDLIHLLGHVPWLLHPSPS 215
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 287 qlrpveagllsakdfiaslafvfgctgyirhaaspmhspepcocchellgvnpladrtfa 346
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 216 EFFINMGRLEFTKIEKVOALPSKQRQTQIQSLAIARCFWFVTESGLIENHEGRKAYG 275
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 347 qfsgdig-----laslgasdeeklst-----vywftvefglckqngelkayg 390
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 276 AVLSSSQEILGHAFIDNVRVLPLELDQIIRLPENTSPQETLFSIRHF----DELVELTS 331
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 391 agllssygeillssseepevrafpdtaavqpygdqtyqpyvfsvsfndakdklrynas 450
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 332 KLE 334
      :  :  :
Db 451 rig 453

```

[illegible]

FT Modified-site 153
 FT /note= "Phosphorylation site"
 PN US5212082-A.
 XX 18-MAY-1993.
 XX 13-MAR-1991; 91US-0669446.
 XX 13-MAR-1991; 91US-0669446.
 XX (UUNY) UNIV NEW YORK STATE.
 XX Filer D, Friedhoff AJ, Goldstein M, Wu J;
 XX WPI; 1993-175456/21.
 XX Genetically modified tyrosine hydroxylase having increased
 XX activity - used for treating neurological disorders e.g.
 XX Parkinson's and Alzheimer's diseases and affective disorders
 XX Disclosure; Fig 5; 20pp; English.
 CC The sequences given in AAR36740-41 represent the human and rat
 CC tyrosine hydroxylase protein respectively. These sequences may be
 CC used in the construction of a variant tyrosine hydroxylase which
 CC contains at least one amino acid substitution in the N-terminal 55
 CC amino acids. The substitution corresponds to an amino acid selected
 CC from Ser8, Ser31, Arg37, Arg38, Gln39, Ser40, Ile42, Glu43,
 CC Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant
 CC protein is at least 3-fold greater than that for the wild type
 CC protein. Cells transfected with the DNA encoding these proteins may
 CC be used for treating neurological disorders associated with a
 CC deficiency in tyrosine hydroxylase or dopamine. These proteins may
 CC be used to generate antibodies specific for the variant tyrosine
 CC hydroxylases to monitor the enzyme during a treatment regimen.
 XX Sequence 498 AA;
 SQ
 Query Match 11.7%; Score 221; DB 14; Length 498;
 Best Local Similarity 27.4%; Pred. No. 9e-14; Mismatches 36; Gaps 5;
 Matches 65; Conservative 37;
 QY 118 WKS-----YCPREFLDYLEAFGLSLDFLDH-----QAVIKFPELETHPSYVPS 161
 DB 233 wkevvtlkglythacrehlegfqllycgyredsipqlqledvsrflkrtgqlrpva 292
 QY 162 GFVAPHQYLSLQDRYFPPIASVMFTLDKNFSLTPDLIHDLLGHVPWLLHPSPFINM 221
 DB 293 gllsakdfilaslafvfgctqyirhasspmhspepcchellghvpmldrtafqfsqdi 352
 QY 222 GRLETKVIEKVOALPSKKORIOTLOSNIAlVRCFWFTVESGLIENHGRKAYGAVLISS 281
 DB 353 g-----laslgasdeeelek1st-----vwtvfeqlckngelkayagallss 396
 QY 282 POELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLSIRHF-----DELVELTSKLE 334
 DB 397 ygelhlsisepevratpdaaavqpygdqtdyqyvfysesfndakdkirnyasriq 453
 RESULT 4
 ID AAR50184 standard; Protein; 497 AA.
 XX AAR50184;
 AC AAR50184;
 DT 04-NOV-1994 (first entry)
 XX Human tyrosine hydroxylase.
 XX Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;
 KW enhanced enzymatic activity; wild-type; transformation; dopamine;

KW Parkinsons disease; Alzheimers disease; brain; encapsulation;
 XX selectively permeable; polymer capsule; antibody.
 XX Homo sapiens.
 XX US5300436-A.
 XX 05-APR-1994.
 XX 13-MAR-1991; 91US-0669446.
 XX 13-MAR-1991; 91US-0669446.
 XX 26-JAN-1993; 93US-0009075.
 XX (UUNY) UNIV NEW YORK STATE.
 XX Filer D, Friedhoff AJ, Goldstein M, Wu J;
 XX WPI; 1994-125849/15.
 XX N-PSDB; AAQ44699.
 XX Modified DNA encoding variant tyrosine hydroxylase with an
 XX N-terminal amino acid substn. and cells contg. it - can be used
 XX to treat diseases associated with defective function of the
 XX enzyme, e.g. Parkinson's disease or Alzheimer's disease
 XX Disclosure; Column 21-24; 27pp; English.
 XX This sequence represents human tyrosine hydroxylase (TH) and may be
 XX used in the production of the variant TH molecules of the invention.
 XX These variants contain amino acid substitutions in the N-terminal
 XX portion of the molecule, and in the human molecule these
 XX substitutions are pref. at positions Ser19, Ser31, Arg37, Arg38,
 XX Ser40, Leu41, Glu43, Asp44, Ala45, Arg46 or Lys47. Variants
 XX containing one or more of these amino acid substitutions, have
 XX substantially enhanced enzymatic activity compared to the wild-type
 XX enzyme. The most pref. substitution being Ser40 for Tyr or Leu.
 XX DNA encoding the TH variants, and cells transformed with this DNA
 XX may be used for treating diseases associated with defective function
 XX of TH, or dopamine, eg. Parkinson's disease and Alzheimers disease,
 XX and effective disorders. The cells can then be implanted into the
 XX brain or encapsulated in a selectively permeable polymer capsule
 XX which allows release of the cells products but protects them from
 XX attack by the hosts antibodies or cells.
 XX Sequence 497 AA;
 SQ
 Query Match 11.6%; Score 220; DB 15; Length 497;
 Best Local Similarity 26.7%; Pred. No. 1.1e-13;
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;
 QY 112 SSRSFLWKS-----YCPREFLDYLEAFGLSLDFLDH-----QAVIKFPELETHF 155
 DB 226 aesiawkevtlkglythacrehlegfqllycgyredsipqlqledvsrflkrtg 285
 QY 156 SYYPVSGFVAPHQYLSLQDRYFPPIASVMFTLDKNFSLTPDLIHDLLGHVPWLLHPSPFS 215
 DB 286 qlrpvagllsardfiaslafvfgctqyirhasspmhspepcchellghvpmldrta 345
 QY 216 EFFINMGRLETKVIEKVOALPSKKORIOTLOSNIAlVRCFWFTVESGLIENHGRKAYG 275
 DB 346 qfsdqig-----laslgasdeeelek1stis-----wftvfeqlckngelkayg 389
 QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLSIRHF-----DELVELTS 331
 DB 390 agllssygelhlsisepevratpdaaavqpygdqtdyqyvfysesfndakdkirnyas 449
 QY 332 KLE 334
 DB 450 riq 452

3

CC (CT-A- cholera toxin, DT-B/- truncated diphtheria toxin,
CC SITA- Shiga-like toxin A; HIVP-BP- HIV protease binding protein.
CC See also AAQ12710-12.
XX
SQ Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 12; Length 452;
Best Local Similarity 26.6%; Pred. No. 4.5e-13;
Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;
QY 104 RNLMYRLSSRSFLMKSCYCPREFLDYLEAFGLLSDFLDH-----QAVIKFFELETHF 155
DB 184 kktwgtvfklslyktha---cyeynhifllekycgfhednipqledvsgflqtctgf 240
QY 156 SYYPVSGFVAPHQYLSLLQDRYFPPIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSPFS 215
DB 241 rlrpvagllssrdflgglafvfhctgyirhgskpmytpdpdichellghvplfsdrsa 300
QY 216 EFTINMGRLEFTRVIEKVQALPSKKQRIQTQSNLIAIVRCFWFTVESGLIENHGRKAYG 275
DB 301 qfsqelg-----laslgapdeyieklat-----iywftvefglckgqdsikayg 344
QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPENTSTPQETLFSIRHDELVE 328
DB 345 agllssfgelqyclsekpllpielektaiqnyvtvtefplpyyvaesfndake 397

RESULT 7
AAW25788
ID AAW25788 standard; Protein; 452 AA.

AC AAW25788;
XX
DT 27-MAR-1998 (first entry)
DE Human phenylalanine hydroxylase.
KW Phenylalanine hydroxylase; human; hybrid protein; cell delivery;
KW cell binding ligand; translocation domain; diphtheria toxin B';
KW phenylketonuria; therapy.
XX Homo sapiens.
OS
PN US5668255-A.
PD 16-SEP-1997.
PF 07-JUN-1984; 84US-0618199.

27-JUN-1991; 91US-0722484.
PR 07-JUN-1984; 84US-0618199.
PR 25-APR-1985; 85US-0726808.
PR 27-JUN-1984; 84US-0618199.
PR 25-APR-1985; 85US-0722484.
PR 22-DEC-1989; 89US-0456095.
PR 14-JUN-1990; 90US-0538276.
PR 04-AUG-1993; 93US-0102387.
XX (SERA-) SERAGEN INC.
XX
XX Murphy JR;
XX
XX WPI; 1997-470103/43.
XX N-PSDB; AAT91639.
XX
XX New hybrid molecules for delivery of agents to cells - comprise a
XX binding domain of a cell binding ligand and a portion of a
XX translocation domain of a protein
XX
XX Example 5; Fig 13A-C; 30pp; English.
XX
XX This protein comprises for human phenylalanine hydroxylase (PH).
XX A DNA molecule (see AAT91638) encoding PH was used to construct a

CC PH-diphtheria toxin B' gene that was expressed in E. coli. The
CC resulting hybrid protein can be used in the treatment of
CC phenylketonuria. The active PH enzyme is targeted to, and
CC incorporated into, the broad range of cells which native diphtheria
CC normally attacks, achieving the widespread therapy that is needed
CC for this inherited disorder. Claimed hybrid proteins comprise a
CC translocation domain and a cell binding domain. They can be used
CC for the delivery of agents (e.g. therapeutic genes, toxins,
CC detectable labels) into cells. The use of a translocation
CC mechanism ensures that the hybrid will be effective in relatively
CC low doses, since a high proportion of the substance of interest
CC will be taken into the targeted cells. The hybrid molecules can be
CC manufactured as a single hybrid recombinant protein, permitting
CC reproducibility, consistency, and the precise control of
CC composition.
XX
SQ Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 18; Length 452;
Best Local Similarity 26.6%; Pred. No. 4.5e-13;
Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;
QY 104 RNLMYRLSSRSFLMKSCYCPREFLDYLEAFGLLSDFLDH-----QAVIKFFELETHF 155
DB 184 kktwgtvfklslyktha---cyeynhifllekycgfhednipqledvsgflqtctgf 240
QY 156 SYYPVSGFVAPHQYLSLLQDRYFPPIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSPFS 215
DB 241 rlrpvagllssrdflgglafvfhctgyirhgskpmytpdpdichellghvplfsdrsa 300
QY 216 EFTINMGRLEFTRVIEKVQALPSKKQRIQTQSNLIAIVRCFWFTVESGLIENHGRKAYG 275
DB 301 qfsqelg-----laslgapdeyieklat-----iywftvefglckgqdsikayg 344
QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPENTSTPQETLFSIRHDELVE 328
DB 345 agllssfgelqyclsekpllpielektaiqnyvtvtefplpyyvaesfndake 397

RESULT 8
AAW55893
ID AAW55893 standard; Protein; 452 AA.

AC AAW55893;
XX
DT 15-FEB-2000 (first entry)
DE Human phenylalanine hydroxylase.
KW Recombinant; hybrid; binding domain; ligand; animal cell; diphtheria;
KW translocation domain; botulinum; neurotoxin; ricin; cholera; tetanus;
KW shiga-like toxin; pertussis; translocation; cytoplasmic membrane; HIV;
KW cytosol; therapy; genetic deficiency disease; enzyme; co-factor; poison;
XX adipocyte; cancer; virus; infection; antibody.
OS Homo sapiens.
XX
XX US5965406-A.
XX
XX 12-OCT-1999.
XX
XX 07-JUN-1995; 95US-0488246.
XX
XX 04-AUG-1993; 93US-0102387.
XX 07-JUN-1984; 84US-0618199.
XX 27-JUN-1991; 91US-0722484.
XX 25-APR-1985; 85US-0726808.
XX 07-JUN-1985; 85US-0742554.
XX 22-DEC-1989; 89US-0456095.
XX 14-JUN-1990; 90US-0538276.
XX (SERA-) SERAGEN INC.


```
XX SQ Sequence 924 AA;
Query Match 5.2%; Score 99; DB 19; Length 924;
Best Local Similarity 19.9%; Pred. No. 0.46;
Matches 90; Conservative 77; Mismatches 122; Indels 164; Gaps 27;
QY 19 KLRQSLFFQNSQRAYSTPYRYRILOKNEKQALAHKHCISILEFFKNLLFVH 78
DB 233 kvegmtqklen--vlnrasntadtlfgevlgrtkk---adstrnalnvlgrfk-flfnl 286
QY 79 LLSLKNQREGCSDMAVSTPPFNRLNWLRLSSRF-----SLWKSVCPRFFLDYLEAF 133
DB 287 plnlkrnlqkg---dydvind-----yekaksifgktevqvkky-----yaeve 329
QY 134 GLSLDFLDHQAIVIRFFLETHFSYVPVSGFV-----APHQY-LSLLQ 174
330 agiedl--reillklk1-letpstlhdqkryrlysdldhapgdpaqcigagbhkwtlkmq 386
QY 175 D-----RYFFIASVMRTLDKDNFSLTDLIHDILGHV----- 206
DB 387 dckeghmkslkgngpphspm-----ldldn-----darpvlglnstaslkrgssfqg 435
QY 207 ----PWLH-----PSFSEFFINM--GRFTKVIK-----VOALPSK 238
DB 436 grddtwtlykthprvafvekltklvlsqldpnlwkwisvngslfsetaeksgqierknv 495
QY 239 KOR-----IOTLQSNLIAIVR--CFWFTVESGLIENHEGRKAYGAVLIS---SPOELG 286
DB 496 rdgrndfkmiqemhslvklirgallpfsalreg-----dgrq-yggvevqaelsgqqla 549
QY 287 HAFIDNVRV-----LPLELDQIIRLPNTSTPQETLFSIR-----HFDELV 327
DB 550 hv-igtirlyesitaleipndmlqii-----gdilrlrhcmvltlghtaeiek 599
QY 328 ELTSKLEWMLDQLESPLYNQEKYLSGPEVL 360
DB 600 rlaekedwvndnegltalpcqfegsvhslqlsl 632
RESULT 12
AAW19604
XX AAW19604 standard; Protein; 1024 AA.
XX AAW19604;
XX 21-AUG-1997 (first entry)
XX Mycoplasma genitalium 116 kDa protein MG075 useful in vaccine.
XX Mycoplasma; immunogen; vaccine; diagnosis; pneumonia; inflammation.
XX Mycoplasma genitalium.
XX WO9721727-A1.
XX 19-JUN-1997.
XX 13-DEC-1996; 96WO-AU00803.
XX 13-DEC-1995; 95AU-0007127.
XX (UYME ) UNIV MELBOURNE.
XX Brownrig GF, Duffy MF, Walker ID, Whithear KG;
XX WPI; 1997-332722/30.
XX PT. New immunogenic polypeptide(s) from Mycoplasma species - useful in
XX PT vaccines and for diagnosis of Mycoplasma infection
XX Claim 19; Page 85-89; 110pp; English.
```

```
XX CC Isolated or recombinant immunogenic polypeptides from Mycoplasma
XX CC genitalium have mol.wt. of 16 kDa (AAW19603) (MG074) and 116 kDa
XX CC (AAW19604) (MG075). They are homologues of 16 and 116 kDa proteins
XX CC DNA sequence of M. genitalium contains contiguous open reading
XX CC frames that code for the 2 polypeptides. Mycoplasma 16 or 116 kDa
XX CC proteins, or immunogenic fragments that include a T or B cell
XX CC epitope, can be used in vaccines for prevention and treatment of
XX CC Mycoplasma infections, partic. in humans. They can also be used
XX CC diagnostically to detect Mycoplasma, or to raise antibodies useful
XX CC in immunoassays.
XX SQ Sequence 1024 AA;
Query Match 5.2%; Score 98.5; DB 18; Length 1024;
Best Local Similarity 19.9%; Pred. No. 0.6;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;
QY 29 QNSQSQRAYSTPYRYRILOKNEK---QALAHKHCISILEFFKNLLFVHLLSLKN 85
DB 564 qtdslknlfsv----igdlsetcnvknltlhavknnehlisivetastkikh1----- 613
QY 86 QREGCSDMAVSTPPFNRLNWLRLSSRFSLWKSVCPRFFLDYLEAFGLSLDFLDHQA 145
DB 614 -----nvqykviavdkfelknsfik-----ellnffpdktdtpt 647
QY 146 IK--FFELETH-----FSYYPVSGFVAPHQYLSLLQDRYFFPIASVMRTLDKDNF 192
DB 648 ikkvlfesenyktrkkyenegfpgyhwakfivpgtfnsaentfysal-----dkt 698
QY 193 SLTDLIHDILGHVPLHLPSPSEFFINMGRFLT-----KVIEKVAQALPSKKQRIQT 244
DB 699 ksirdlfadml-fgkslesvndsdfikingsftlittkndnlnllpnhyhslitknvgqi 757
QY 245 LOSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSSPOELG----- 286
DB 758 vvnvfhidarlltaeqntvfsnpk-----pvikspvelskslfwwktifensvni 810
QY 287 ---HAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFD 324
DB 811 lkkeytfknlkffpkadgsrleldskpdrvipfafvd 852
RESULT 13
AAW19604
XX AAW19604 standard; Protein; 439 AA.
XX AAW19604;
XX 13-SEP-1999 (first entry)
XX C. pneumoniae protein involved in metabolism of nucleic acids.
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX vaccine; neutralising epitope.
XX Chlamydia pneumoniae.
XX WO9927105-A2.
XX 03-JUN-1999.
XX 20-NOV-1998; 98WO-IB01890.
XX 04-NOV-1998; 98US-0107078.
XX 21-NOV-1997; 97FR-0014673.
XX (GEST ) GENSET.
XX Griffais R;
```


PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 17-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.0%; Score 95; DB 21; Length 920;
Best Local Similarity 21.7%; Pred No. 1, 2; Mismatches 174; Gaps 23;
Matches 97; Conservative 48;

QY 9 DPXYILKIALKRLQSLSLFQNSOSLORA-----YSTPYSYRILLOKENEKQALAR 61
Db 371 npaivqkvdelspnsfrifwesqfegtdkaepwyncayslekfss-----419
QY 62 HKCISILEFFKNLLFVHL-----LSLSKNQREGCSTDMVSTPFFNRLWYR 109
Db 420 ----tlqewvqsapdvhlhlpapnvpfiptdls-lkdaadketvpvlrktpf--srlwyk 472
QY 110 LLSRFSLMKSY-----Cp-----REFLDYLEAF-----133
Db 473 -pdtmfakpkayvkmfncplavsspdavltidfrlmdylneyayyaqvaglyygs 531
QY 134 -----GLLSDF---LDHOAVIKFFFELETHFSYYPVSGFA 165
Db 532 lsdngfeltllgynhkirillettvgkianfevkpdrfavik---etvkeyqnykfrq 587
QY 166 p-HQ---YLSL-LQDRYFPIASVMRTLDKNFSLTPDLIHLLGH-----VPWLLH 211
Db 588 pyhqmyyccslilqdtwp-----wteel--dvlsheadevkvfpmlls 631
QY 212 PSSEPEI-----NMGRUFTKVIKV-----QALPSKKQRIOTLOSMLIAIVRC 255
Db 632 rtflecyiagnvenneasmvkhiedvfnpdkpicrplfs-----ghlnrvvki---683
QY 256 FWTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVRLPDLLEDOIIRLP-FNTSTPQ 314
Db 684 -----gegmkfyhqdgnsdensalvhyiqvhrddfsmnkiklqifglvakq 731
QY 315 ETLSIRHFDELVELTSKLEWMLDOGL 341
Db 732 atfhqtrtveqlgyita-laqrndagi 757

RESULT 15

AAG31678
ID AAG31678 standard; Protein; 944 AA.
XX AAG31678;
AC AAG31678;
XX AAG31678;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 38083.
XX Arabidopsis thaliana.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 16-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134321.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 24-MAY-1999; 99US-0135353.
XX 25-MAY-1999; 99US-0135629.
XX 27-MAY-1999; 99US-0136021.
XX 28-MAY-1999; 99US-0136392.
XX 01-JUN-1999; 99US-0136782.
XX 03-JUN-1999; 99US-0137222.
XX 04-JUN-1999; 99US-0137328.
XX 07-JUN-1999; 99US-0137502.
XX 08-JUN-1999; 99US-0137724.
XX 10-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 14-JUN-1999; 99US-0138847.
XX 16-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140961.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

```
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151203.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 5.0%; Score 95; DB 21; Length 944;
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;

QY 9 DPKYILKALRQSLSLFQNSQLQRA-----YSTPYSYVYRIILQKENKEKQALAR 61
Db 395 npaivqkvvdelspsnfrifwesqkfegqtdkaepwyntayslekitts----- 443
QY 62 HKCISILEFFKNLLFVHL-----LSLSKNOREGCTDMAVVSTPFENRNLWYR 109
Db 444 ----tiqewqsapdvhlhlpapnfvipdtsl-kdaadketvpyllrktpf--srlwyk 496
QY 110 LLSRFSFLWKSX-----CP-----RFLDYLEAF----- 133
Db 497 -pdtmfskpavkmdncplavsspsdaavltidfrllmdylneyayyaqvaylygvs 555
QY 134 -----GLLSDF-----LDHQAVIKFEELETHFSYYPVSGEVA 165
```

```
Db 556 lsdngfeltllgynhklrillevvvgkianfevkpdrfavik-----etvtkeygnykfrq 611
QY 166 P-HQ---YLSL-LQDRYFPIASVMRTLDKDNFSLTPDLIHLLGH-----VPWLLH 211
Db 612 pyhqamyycslilqdtwp-----wteel--dvishleaedvakfvpmlis 655
QY 212 PSSEFFI-----NMGRFTKVIKV-----QALPSKKQRIQTLOSNLIAIVRC 255
Db 656 rtflecyiagnvenneaesmvkhiedvifndpkpicrplfps-----ghltnrvvki--- 707
QY 256 FWTVESGLIENHEGRKAYCAVLISSPQELGHAFIDNVVRLPLELDQIIRLP-FNTSTPQ 314
Db 708 -----gegmkyfyhqdgnspsdensalvhyiqvhrddfsmniklqifglvakq 755
QY 315 ETLSIRHFDVELTFSKLEWMLDQGL 341
Db 756 atfhqlrtveqlgyita-laqrndsgi 781
```

Search completed: June 29, 2001, 13:52:48
Job time: 7960 sec


```
34 euGlnArgAlaTyrSerThrProTyrSerTyrTyrArgIleIleLeuGln 50
|||||
1210726TCCACAGTGCATACGACCCCATATCTCTACTACCGAATCATCTACAA 1210677
|||||
51 LysGluAsnLysClnLysClnAlaLeuAlaArgHisLysCysIleSerIle 67
|||||
1210676AAGGAAATTAAGAGAGACGACCTTACGTCGACACAAATGCATTTCTAT 1210627
|||||
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
|||||
1210626TTTGTAGAAATTTTCAAAACTTACTCTTTGTTTCATCTCTCTGCATTTATCAA 1210577
|||||
84 ysAsnGlnArgGluGlyCysSerThrAspMetAlaValValSerThrPro 100
|||||
1210576AGAAATCAAAAGGGAAGGTGCTCCACATGATATGGCTGTTGTAAGCACTCCC 1210527
|||||
101 PheAsnArgAsnLeuTyrPtyrArgLeuLeuSerSerArgPheSerLe 117
|||||
120526TTTTTAATCGGAATTTATGATGATGACCTCTTTCCTCAGGGTTTCTCT 1210477
|||||
117 uTrpLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
|||||
1210476ATGGAAGCATTTGTCCAAGATTTTCTT.GATTACTTAGAAGCTTTCG 1210428
|||||
134 LysLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPheGlu 150
|||||
1210427GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1210378
|||||
151 LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProH 167
|||||
1210377TTAGAAACACATTTTTCATTTATTCCTGTTTCAGGATTTGATCTCCCA 1210328
|||||
167 sGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerVal 184
|||||
1210327TCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1210278
|||||
184 etArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHis 200
|||||
1210277TGGGAATCTCGATAAAGATAATTTCTCTTAACTCTGATCTCATCCAT 1210228
|||||
201 AspLeuLeuGlyHisValProTrpLeuLeuHisProSerPheSerGluPh 217
|||||
1210227GACCTTTTAGGGCAGCTGCTTGGCTTCTACATCCCTCATTTCTGAAT 1210178
|||||
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGln 234
|||||
1210177TTTCATAAATGGAAGACTCTTCACTAAAGTCATAGAAAAAGTACAAG 1210128
|||||
234 laLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
|||||
1210127CTCTCTAGTAAAAACAACGACATACAAACCTTACAAAGCATCTGATC 1210078
|||||
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267
|||||
1210077GCTATGTACGCTGCTTTGGTTTACTGTTGAAGCGGATTTATGAA 1210028
|||||
267 nHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnG 284
|||||
1210027CCATGAAGGAAGAAAGCATATGAGCGGCTTATCAGTCTCCACAG 1209978
|||||
284 luLeuGlyHisAlaPheLeuAspAsnValArgValLeuProLeuGluLeu 300
|||||
1209977AACTTGACACGCTTCTATGTATGACGTCAGTGTCTCCCTTTAGAAATG 1209928
|||||
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLe 317
|||||
1209927GATCAGATTTATGCTCTTCCCTTCAATATCATCACTCCACAAGAGACT 1209878
|||||
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334
|||||
1209877ATTTTCAATAAGACATTTTGTATGAACCTGTAAGACTCACTTCAAAATAG 1209828
|||||
```

```
334 lutrMetLeuAspGlnGlyLeuLeuGluSerIleProLeuTyrAsnGln 350
|||||
1209827AATGGATGCTCGACCAAGGCTGTTAGTAATCAATCCCTTTTACAATCAA 1209778
|||||
```

```
351 GluLysTyrLeuSerGlyPheGluValLeuCysGln 362
|||||
1209777GAGAATATATCTTCTGGTTTGGAGTACTTTGGCCAA 1209742
|||||
```

seq_name: /SIDS1/gcdata/geneseq/geneseq/NA1997.DAT:AAAT62529

seq_documentation_block:

ID AAAT62529 standard; DNA; 1030 BP.

XX AC AAAT62529;

XX DT 06-JUN-1997 (first entry)

XX DE Tyrosine hydroxylase truncated DNA construct rTHdel.

XX KW Analgesic; pain; bioartificial organ; tyrosine hydroxylase;

XX KW norepinephrine; rTHdel; catecholamine; ss.

XX OS Rattus sp.

XX PH Key Location/Qualifiers

XX FT 1..6 1..6

XX FT 5'UTR /*tag= a

XX FT exon 1..1030

XX FT /*tag= b

XX FT /codon_start= 7

XX FT 1018..1030

XX FT /*tag= c

XX PN WO9640959-A1.

XX PD 19-DEC-1996.

XX PF 07-JUN-1996; 96WO-US09629.

XX PR 07-JUN-1995; 95US-0481917.

XX PA (CYTO-) CYTOTHERAPEUTICS INC.

XX PI Saydoff J, Wong S;

XX DR WPI; 1997-087062/08.

XX PT Stably transformed cells expressing endorphin, enkephalin and

XX PT catecholamine - and artificial organs contg. them, useful for

XX PT control of pain, esp. implanted in the CNS

XX PS Example; Page 66; 114pp; English.

XX CC A DNA sequence (AAAT62529), designated rTHdel, codes for truncated rat

XX CC tyrosine hydroxylase. It was generated by PCR amplification (see

XX CC also AAAT62526-27) of TH cDNA and insertion of the PCR product into

XX CC vector pCDNA3. rTHdel can be utilized in the development of

XX CC truncated TH in e.g. rat pancreatic endocrine KIN host cells.

XX CC Such cells, pref. also contg. dopamine beta-hydroxylase sequences,

XX CC are capable of norepinephrine synthesis. Sequential transfection

XX CC allows produ. of cell lines that produce more than one analgesic

XX CC cpd. Such cell lines can be encapsulated to form bioartificial

XX CC organs that can be implanted e.g. in the CNS for the control of

XX CC pain.

XX SQ Sequence 1030 BP; 223 A; 301 C; 274 G; 232 T; 0 other;

alignment_scores:

Quality: 223.00

Length: 243

Gaps: 5

Ratio: 1.640


```

1273 CGCGAAGAGATTGCTACCTCGGAGGAGGTATATGTCACGCTGAAGGCGCT 1322
121   TTrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuS 137
1323 CTATGCTACCATGCCCTCGGAGGACCTGAGGGTTTCAGCTTCGG 1372
137   eRaspPheLeuAspHis.....GlnAlaVal 145
1373 AACGTGATGGCTACCGGAGGAGACAGCATCCACAGCTGGAGGAGCTG 1422
146   lLeLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerGl 162
1423 TCCCGCTCTTTGAAGGAGCGAGCTGGCTTCCAGCTGCGACCCGTCGCCG 1472
162   yPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheP 179
1473 TCTACTGTCGCCCGTGATTTCTGCCAGTCTGCCCTTCGCGTTCGCGTGTTC 1522
179   rOileAlaSerValMetArgThrLeuAspLysAspAsnPheserLeuThr 195
1523 AATGCACCAGTATATCCGCACTGCTCCTCACCTATGCATTACACCTGAG 1572
196   PrAspLeuIleHisAspLeuLeuGlyHisValProTPrLeuLeuHisPr 212
1573 CGGAGCTGCTGCCATGAGCTGTGTGGACATGTATACCATGTTGGCTGACCG 1622
212   oSerPheSerGluPheIleAlaMetGlyArgLeuPheThrLysValI 229
1623 CACATTTGCCAGTCTCCAGGACATTCGA..... 1653
229   leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
1654 .....CTTGCACTCTCGGGGCTCTCAGATGAAGAAATGAAAACTC 1695
246   GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262
1696 TCCAGC.....GTGTACTGGTTCATCTGCGGAATT 1724
262   rGlyLeuLeuGluAsnHisGluGlyArgLysLysAlaTyrGlyAlaValLeuI 279
1725 CGGGCTATGTAACACAGAAATGGGAGCTGAAGCTTATCGTCGAGGGCTGC 1774
279   leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
1775 TGTCTTCTCAGGAGAGCTCTCATCTCCCTCTCAGAGGAGCGCTGAGGTC 1824
296   LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
1825 CGAGCGCTTTGTACCAGACACAGCAGCTGTGCAGCGCCTACCAAGATCAAC 1874
312   rProGlnGluThrLeuPheSerIleArgHisPhe.....AspG 325
1875 CTACAGCGCTGTGTAATTGTGTCCGAGAGCTTCAATCAGCCCAAGGACA 1924
325   lueValGluLeuThrSerLysLeuGlu 334
1925 AGCTCAGGACACTATGCCTCTCGTATCCAG 1953

seq name: /SIDS1/gcgdata/geneseq/geneseq/Nal97.DAT.AAT62548
seq documentation_block:
ID   AAT62548 standard; DNA; 5540 BP.
XX
AC   AAT62548;
XX
DT   07-JUN-1997 (first entry)
XX
DE   IgSP-hPOMCdeltaACTH-IRES-rTHdel-IRES-bDBH-IRES-zeocin-073 DNA.
XX
KW   Analgesic; pain; bioactive; functional organ; pro-opiomelanocortin; POMC;
KW   beta-endorphin; tyrosine hydroxylase; dopamine beta-hydroxylase;
KW   IgSP-hPOMC-deltaACTH-IRES-rTHdel-IRES-bDBH-zeocin-073;
KW   internal ribosome entry site; ss.

```



```

112 SerSerArgPheSerLeuTrpLysSer..... 120
199 GCGAGAGATGCTACCTGGAAGGATATATGTCACGCTGAAGGCCT 248
121 .TyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuS 137
249 CTATGCTACCATCCATCGCTGCCGGGACACCTGGAGGTTTCCAGCTTCGG 298
137 erAspPheLeuAspHis.....GlnAlaVal 145
299 AACGGTACTGCTGCTACCGAGAGGACACATCCACAGCTGGAGGACGTG 348
146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerG 162
349 TCCGGCTTCTGAAGGAGCGGACTGCTCCAGCTGGACCGTGGCGG 398
162 yPheValAlaProHisGlnTyrLeuSerLeuGlnAspArgTyrPheP 179
399 TCTACTGTCGGCCGATGATTTCTGGCCAGCTGCGCTCCGCTGTTTC 448
179 rolleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
449 AATGCCACCCATATATCCGCAATGCTCTCCTCACCCTATGATTCACCTGAG 498
196 ProAspLeuIleHisAspLeuGlyHisValProTyrLeuLeuHisPr 212
499 CCGGACTGTCGCATGAGTGTGGACATGTACCCATGTTGGCTGACCG 548
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
549 CACATTTGCCAGTTCTCCAGGACATTTGA..... 579
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
580 .....CTTGCACTCTGGGGCCCTCAGATGAAGAAATGAANAACTC 621
246 GluSerAsnLeuIleAlaIleValArgCysPheThrPheThrValGluSe 262
622 TCCAG.....GTGTACTGGTTCACCTGTGGAAT 650
262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeu 279
651 CGGCTATGTACAGAAATGGGAGCTGAGGCTTATGTGTCAGGCTGC 700
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
701 TGTCTTCTACGAGAGCTCTGCACTCCCTGTGCAGAGGAGCTGAGGTC 750
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
751 CGAGCCTTTGACCCAGACACAGACAGCTGTGCAGCCCTTACCAAGATCAAC 800
312 rProGlnGluThrLeuPheSerIleArgHisPheAspGlu 325
801 CTACACGCTGTGTACTTGTGTCGAGAGCTTCAATGAC 840

seq_name: /SIDS1/gcgdata/geneseq/NA1997.DAT:AA62536
seq_documentation_block:
ID AA62536 standard; DNA; 3432 BP.
XX
AC AA62536;
XX
DT 06-JUN-1997 (first entry)
XX
DE rThdelKS-IRES-bDBH DNA sequence.
XX
KW Analgesic; pain; bioartificial organ; tyrosine hydroxylase;
KW dopamine beta-hydroxylase; internal ribosome entry site; IRES;
KW norepinephrine; catecholamine; rThdelKS-IRES-bDBH; ss.
XX
OS Chimeric Rattus sp.;
XX Chimeric picornavirus;

```

```

OS Chimeric Bos taurus.
XX Key Location/Qualifiers
FH 5'UTR 1..13
FT exon /*tag= a
FT exon 1..1024
FT intron /*tag= b
FT intron 1025..1624
FT exon /*tag= c
FT exon 1625..3432
FT 3'UTR /*tag= d
FT misc_feature 3419..3432
FT /*tag= e
FT /*tag= f
FT /*product= IRES
XX WO9640959-Al.
XX 19-DEC-1996.
XX 07-JUN-1996; 96WO-US09629.
XX 07-JUN-1995; 95US-0481917.
XX (CYTO-) CYTOTHERAPEUTICS INC.
XX Saydoff J, Wong S;
XX WPI; 1997-087062/08.
XX Stably transformed cells expressing endorphin, enkephalin and
XX catecholamine - and artificial organs contg. them, useful for
XX control of pain, esp. implanted in the CNS
XX Example; Page 72-75; 114pp; English.
XX 2 DNA constructs (AA62535 and AA62536) respectively comprise a
XX truncated rat tyrosine hydroxylase sequence, rThdel (see also
XX AA62529) or rThdelKS (see also AA62530), joined via an IRES sequence
XX to the bovine dopamine beta-hydroxylase gene. Expression of the
XX constructs in transfected RIN or AtT-20 cells complements the
XX host cells' catecholamine synthesising enzymes, allowing prodn.
XX of norepinephrine. Sequential transfection of host cells with
XX diff. vectors, or with a polycistronic vector (see also AA62543,
XX AA62548), allows prodn. of cell lines that produce more than one
XX analgesic cpd. Such cell lines can be encapsulated to form
XX bioartificial organs that can be implanted e.g. in the CNS for the
XX control of pain.
XX SQ Sequence 3432 BP; 688 A; 1098 C; 963 G; 683 T; 0 other;

alignment_scores:
Quality: 221.00 Length: 230
Ratio: 1.713 Gaps: 4
Percent Similarity: 56.087 Percent Identity: 26.957

alignment_block:
US-09-438-185-1047 x AA62536 ..
Align seg 1/1 to: AA62536 from: 1 to: 3432
112 SerSerArgPheSerLeuTrpLysSer..... 120
206 GCGAAGAGATGCTACCTGGAAGGATATATGTCACGCTGAAGGCCT 255
121 .TyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuS 137
256 CTATGCTACCATCCATCGCTGCCGGGACACCTGGAGGTTTCCAGCTTCGG 305
137 erAspPheLeuAspHis.....GlnAlaVal 145

```


945 AGTGACCCAGTATATATCGCCACGCGTCTCGCCCATGCACTCCCTCGAG 994
196 ProAspLeuIleHisAspLeuGlyHisValProTrpLeuHisPr 212
995 CCGAGCTGTGCGACGAGCTGCTGGGACAGTCCCATGCTGCGCCAGCG 1044
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
1045 CACCTTCGGCGAGTTCCTCGAGGACATGGC.....C 1076
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
1077 TGGCGTCTGCTGGGCGCTCGGATGAGGAATGAGAGCTGTCCAGCTG 1126
246 GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGlu 262
1127 TCA.....TGGTTCACGCTGAGATT 1146
262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyGlyAlaValLeuI 279
1147 CGGCGTGTGTAAGCAAGACGGGAGGTGAAGGCTATGTGCGGCGTGC 1196
279 leSerSerProGlnLeuLeuGlyHisAlaPheIleAspAsnValArgVal 295
1197 TGTCTCTACGCGGAGCTCTGCGCATGCTGTGTGAGGAGCTGAGATT 1246
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
1247 CGGCGCTTCGACCTGAGCTGCGCGCTGCGCCCTACCAAGACGAGAC 1296
312 rProGlnThrLeuPheSerIleArgHisPhe.....Asp 325
1297 GTACCACTAGTCTACTCTGCTGTGAGAGTTCAGTTCAGTCCAGCCAGGAC 1346
325 leuLeuValGluThrSerLysLeuGlu 334
1347 AGCTCAGGAGCTATGCTCAGCAGCCAGCCAG 1375

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ90020

seq_documentation_block:

ID AAZ90020 standard; cDNA; 1650 BP.

XX AC AAZ90020;

XX 05-MAY-2000 (first entry)

XX Human phenylalanine hydroxylase cDNA sequence.

XX Phenylalanine hydroxylase; translocation domain; cell destruction;
XX cell binding domain; genetic deficiency disease; cancer;
XX adipocyte; enzyme delivery; anti-viral; HIV; ss.

XX OS Homo sapiens.

XX US6022950-A.

XX 08-FEB-2000.

XX 07-JUN-1995; 95US-0479510.

XX 07-JUN-1984; 84US-0618199.

XX 27-JUN-1991; 91US-0722484.

XX 25-APR-1985; 85US-0726808.

XX 07-JUN-1985; 85US-0742554.

XX 22-DEC-1989; 89US-0456095.

XX 14-JUN-1990; 90US-0538276.

XX 04-AUG-1993; 93US-0102387.

XX (SERA-) SERAGEN INC.

XX Murphy JR;

DR WPI: 2000-160390/14.
DR P-PSDB; AAY78593.

XX New two-part hybrid protein comprising a translocation domain and a
PT cell-binding domain, for treating genetic deficiency diseases, cancer
PT and HIV infections -

XX Example 5; Fig 13; 32pp; English.

XX This sequence represents the phenylalanine hydroxylase coding sequence.
CC The encoded protein can be included in the hybrid protein of the
CC invention and used to destroy or modify the cell that the hybrid protein
CC is targeted to. The hybrid protein comprises a first part which is a
CC portion of the binding domain of a cell-binding ligand, effective to
CC cause the hybrid molecule to bind to a cell of an animal. The second
CC part comprises a portion of a translocation domain of a naturally
CC occurring protein (e.g. the translocation domain of diphtheria toxin) the
CC second part translocates the third part across the cytoplasmic membrane
CC and into the cytosol of the cell. The third part comprises a chemical
CC entity to be introduced into the cell, where each of the first and third
CC part is non-native with respect to naturally occurring protein, and the
CC covalent bond attaching the second and third part is cleavable. The
CC phenylalanine hydroxylase protein can form part of the third portion of
CC the hybrid protein. The cell binding domain binds to a specific cell and
CC the translocation domain transfers the hybrid molecule across the cell
CC membrane into the cytosol. The third part of the protein, linked to the
CC translocation domain through a cleavable bond, can then carry out its
CC function. The hybrid molecules are useful for treating genetic deficiency
CC diseases by delivering to affected cells an enzyme supplying the missing
CC function, to supplement cellular levels of a particular enzyme or a
CC scarce precursor or cofactor, to direct toxins or other poisons to
CC destroy particular cells (such as adipocytes, cancer cells, or
CC virus-infected cells), and to counteract viral infections such as HIV by
CC introducing into appropriate cells antibodies to viral proteins.

XX Sequence 1650 BP; 433 A; 405 C; 386 G; 426 T; 0 other;

alignment_scores:

Quality: 213.50 Length: 233

Ratio: 1.642 Gaps: 4

Percent Similarity: 55.794 Percent Identity: 26.609

alignment_block:

US-09-438-185-1047 x AAZ90020 ..

Align seg 1/1 to: AAZ90020 from: 1 to: 1650

104 ArgAsnLeuTrpTrpArgLeuLeuSerSerArgPheSerLeuTrpLys 120

772 AAGAAACATGGGCGACAGTGTCAAGACTCTGAAGTCTTGATATAAAC 821

120 rTyrcysProArgPhePheLeuAspTyLeuGluAlaPheGlyLeu 137

822 CCATGCT.....TGCTATGAGTACAATCACTTTTCCACTCTTCG 862

137 erAspPheLeuAspHis.....GlnAlaVal 145

863 AAAAGTACTGTGCTTCCATGAGATACATTCGCCAGCTGGAGACGTT 912

146 IleLysPhePheGluLeuGluThrHisPheSerTyTrpProValSerG 162

913 TCTCAATTCCTCGAGACTTGCACCTGTTTCCGCCCTCCGACCTGTGGCTG 962

162 yPheValAlaProHisGlnTyLeuSerLeuGlnAspArgTyPheP 179

963 CCGCTTCTCTCGGAGTTTCTGGTGGCTGCGCTTCCGAGCTCTTC 1012

179 rolleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeu 195

1013 ACTGCACACAGTACATCAGACATGGATCCCAAGCCCATGTATACCCCGAA 1062

196 ProAspLeuIleHisAspLeuLeuGlyHisValProTrpLeuHisPr 212

246 GlnSerAsnLeuIleAlaIleValArgCysPheThrValGlu 262
 1186 GCCACA.....ATTACTGTTTACTGTGGAGTT 1214
 262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValGlu 279
 1215 TGGGCTCTCAACAAGAGACTCCATAAGGATATGCTGGTGGCTCC 1264
 279 leSerSerProGlnLeuGlyHisAlaPheIleAspAsnValArgVal 295
 1265 TGTATCCTTGGTGAATACAGTACGTTATCAGAGAGGCAAGGTT 1314
 296 LeuProGlnLeuAspGlnIleAlaArgLeuPropheAsnThrSerTh 312
 1315 CTCCCTGGAGCTGGAGAGACAGCCATCCAAATACACTGTCCACGA 1364
 312 rProGlnGluThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328
 1365 GTTCCAGCCCTGTATTACGGTCAGAGAGTTTAAATGATGCCAAGGAG 1413

seq name: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ30664

Documentation_block:

AAZ30664 standard; DNA; 2448 BP.

AAZ30664;

15-FEB-2000 (first entry)

Human phenylalanine hydroxylase gene.

Recombinant; hybrid; binding domain; ligand; animal cell; diphtheria;
 translocation domain; botulinum; neurotoxin; ricin; cholera; tetanus;
 shiga-like toxin; pertussis; translocation; cytoplasmic membrane; HIV;
 cytosol; therapy; genetic deficiency disease; enzyme; co-factor; poison;
 adipocyte; cancer; virus; infection; antibody; ss.

Homo sapiens.

US5965406-A.

12-OCT-1999.

07-JUN-1995; 95US-0488246.

04-AUG-1993; 93US-0102387.

07-JUN-1984; 84US-0618199.

27-JUN-1991; 91US-0722484.

25-APR-1985; 85US-0726808.

07-JUN-1985; 85US-0742554.

22-DEC-1989; 89US-0456095.

14-JUN-1990; 90US-0538276.

(SERA-) SERAGEN INC.

Murphy JR;

WPI: 1999-632431/54.

P-PSDB; AAY55893.

Recombinant DNA molecule encoding a three part hybrid protein used in
 the treatment of Aids and genetic deficiency diseases -
 Example 5; Fig 13; 3lpp; English.
 The invention relates to a recombinant DNA molecule encoding a hybrid
 protein comprising three parts: (a) the first part comprises a portion
 of the binding domain of a cell-binding polypeptide ligand allowing the
 hybrid protein to bind to an animal cell; (b) the second part comprises
 a portion of a translocation domain of a naturally occurring protein
 selected from diphtheria toxin, botulinum neurotoxin, ricin, cholera
 toxin, LT toxin, C3 toxin, shiga-like toxin, pertussis toxin and tetanus
 toxin, which translocate the third part of the across the cytoplasmic

CC membrane into the cytosol of the cell; and (c) the third part comprises
 a polypeptide entity to be introduced into the cell, which is non-native
 to the naturally occurring protein of (b). This sequence represents the
 human phenylalanine hydroxylase gene for use in generating the hybrid of
 the invention. The hybrid molecule enables the direction of appropriate
 therapy to affected cells, allowing them to function properly and
 alleviate or cure the disease. The hybrid is especially used in treating
 genetic deficiency diseases, by delivering to affected cells an enzyme
 supplying the missing function, to supplementing cellular levels of a
 particular enzyme or a scarce precursor or cofactor, to directing toxins
 or other poisons to destroy particular cells (such as adipocytes, cancer
 cell, or virus infected-cells), to counteracting viral infections, such as
 HIV, by introducing appropriate antibodies to viral proteins. It is also
 involved in the process of getting non-therapeutic substances such as
 CC detectable labels into cells.

XX
 SQ Sequence 2448 BP; 739 A; 514 C; 501 G; 694 T; 0 other;

alignment_scores:

Quality: 213.50 Length: 233

Ratio: 1.642 Gaps: 4

Percent Similarity: 55.794 Percent Identity: 26.609

alignment_block:

US-09-438-185-1047 x AAZ30664

Align seg 1/1 to: AAZ30664 from: 1 to: 2448

104 ArgAsnLeuTrpTyrArgLeuSerSerArgPheSerLeuTrpLysSe 120
 772 AAGAAACATGGGCACAGCTGTCAGACTCTGAAGCTCTGTATATAAC 821
 120 rTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeu 137
 822 CCATGCT.....TGATAGTACATACATCATTTTCCACACTCTTG 862
 137 erAspPheLeuAspHis.....GlnAlaVal 145
 863 AAAAGTACTGTGGCTTCCATGAAGATAACATCCCGAGCTGGAAGCGTT 912
 146 IleLysPhePheGluLeuGluThrHisPheSerTyrTrpValSerG 162
 913 TCTCAATTCCTGCAGACTTGCACCTGGTTCCGCGCTCCGACCTGTGGCTG 962
 162 yPheValAlaProHisGlnTyrLeuSerLeuGlnAspArgTyrPheP 179
 963 CTGCTTTCCTCTCGGGATTCTTGGTGGCTGGCTTCCGAGCTCTCC 1012
 179 rIleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
 1013 ACTGCACACAGTACATCAGACATGGATCCAAAGCCCATGTATACCCCGAA 1062
 196 ProAspLeuIleHisAspLeuGlyHisValProThrLeuLeuHisPr 212
 1063 CTGACATCTGCCATGAGCTGTGGACATGTGCCCTGTGTTTCAGATCG 1112
 212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysVal 229
 1113 CAGTTTGGCCAGTTTCCAGGAATTTGGC..... 1143
 229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
 1144CTTGCTCTCTGGGTGCACCTGATGATACATGATAAAGCTC 1185
 246 GlnSerAsnLeuIleAlaIleValArgCysPheThrValGluSe 262
 1186 GCCACA.....ATTACTGTTTACTGTGGAGTT 1214
 262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValGlu 279
 1215 TGGGCTCTCAACAAGAGACTCCATAAGGATATGCTGGTGGCTCC 1264

279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
1265 TGTATCCCTTTGGTGAATACAGTACTGCTATCATCAGAGAACCAAGCTT 1314
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
1315 CTCCCTCGAGCTGGAGAGACAGCCATCCAAATATACACTGTACAGGA 1364
312 rProGlnGluThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328
1365 GTTCCAGCCCTGTATTACGTGGCAGAGAGTTTAAATGATGCCAAGGAG 1413
seq_name: /SIDS1/scgdata/geneseq/geneseq/NA1991.DAT:AAQ12712
seq_documentation_block:
ID AAQ12712 standard; DNA: 2523 BP.
XX AAQ12712;
XX 08-OCT-1991 (first entry)
XX Phenylalanine hydroxylase.
XX Hybrid; fusion; membrane translocation; binding region; HIV;
XX infection; toxin; steroid; hormone; monoclonal antibody; antigen;
XX diphtheria; exotoxin; phenylketonuria; cholera; interleukin; IL-2;
XX protease; epidermal growth factor; ricin; tetanus; hexosaminidase;
XX Shiga-like toxin A; SLT-A; PH; ligand; insulin; nuclease; ss.
XX Vibrio cholera.
XX Key Location/Qualifiers
XX CDS 223..1578
XX /*tag= a
XX /label= PH
XX
XX WO9109871-A.
XX 11-JUL-1991.
XX
XX 21-DEC-1990; 90WO-US07619.
XX
XX 14-JUN-1990; 90US-0538276.
XX 22-DEC-1989; 89US-0456095.
XX (SERA-) SERAGEN INC.
XX Murphy JR;
XX WPI: 1991-222845/30.
XX P-PSDB; AAR13119.
XX Hybrid molecules for targeting chemical entity to cell - have
XX membrane trans-locating and cell binding-regions and used to
XX treat HIV infection, genetic enzyme-deficiency disorders etc.
XX
XX Disclosure; Fig 13(1-3); 59pp; English.
XX
XX Hybrid molecules are produced by covalently linking
XX (1) a portion (A) of the binding domain of a cell-binding ligand,
XX allowing binding of the mol. to an animal cell;
XX (2) a portion (B) of a translocation domain of a protein able to
XX translocate (C) across the cell cytoplasmic membrane, and
XX (3) a portion (C) which is to be introduced into the cell.
XX (A) is derived from a steroid or polypeptide hormone, a single-chain
XX analogue of a monoclonal antibody able to bind an antigen expressed
XX on the cell surface, or a polypeptide toxin.
XX (B) is derived from a toxin (e.g. diphtheria toxin or Pseudomonas
XX exotoxin A).
XX (A) may be derived from insulin, interleukins 2, 3 or 6 or
XX epidermal growth factor.
XX Suitable enzymes in (C) include cholera toxin, ricin, tetanus toxin,
XX hexosaminidase A, protease, nuclease, SLT-A, etc.

CC Specified examples are CT-A/DT-B'/IL-2, SUTA/DT-B'/IL-2,
CC ricin A/DT-B'/IL-2, HIVP-BP/DT-B'/IL-2 and the phenylalanine
CC hydroxylase-DT-B' or their biologically active mutants.
CC (CT-A- cholera toxin, DT-B'- truncated diphtheria toxin,
CC SUTA- Shiga-like toxin A; HIVP-BP- HIV protease binding protein.
CC See also AAQ12710-12.
XX
XX Sequence 2523 BP; 766 A; 522 C; 518 G; 717 T; 0 other;
alignment_scores:
Quality: 213.50 Length: 233
Ratio: 1.642 Gaps: 4
Percent Similarity: 55.794 Percent Identity: 26.609
alignment_block:
US-09-438-185-1047 x AAQ12712
Align seg 1/1 to: AAQ12712 from: 1 to: 2523
104 ArgAsnLeuTrpTyrArgLeuLeuSerSerArgPheSerLeuTrpLysSe 120
772 AGAACAATGGCCACAGTGTCAAGACTCTGAAGTCTCTGTATTAAC 821
120 rTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuS 137
822 CCATGCT.....TGCTATGAGTACATCACAATTTTCCACTCTGTG 862
137 erAspPheLeuAspHis.....GlnAlaVal 145
863 AAAAGTACTGTGGCTTCCATGAAGATAACATTTCCCGAGCTGGAGAGCTT 912
146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerGl 162
913 TCTCAATTCCTGCGACTTGCACCTGTTTCCGCTCCGACCTGTGGCTGG 962
162 yPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheP 179
963 CTGTCTTCTCTCGGATTTCTTGGTGGCTTGGCTTCCGAGCTTCC 1012
179 toIleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
1013 ACTGCACACAGTACATCAGACATGGATCCCAAGCCCATGTATACCCCGAA 1062
196 ProAspLeuIleHisAspLeuLeuGlyHisValProTrpLeuLeuHisPr 212
1063 CCTGACATCTCCATGAGCTGTGGACATGTGCCCTGTTTTCAGATCG 1112
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
1113 CAGCTTCCCGAGTTTCCCGAGAAATGEC..... 1143
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
1144CTTGGCTCTCTGGGTGCACCTGATGATACATTGAAAGCTC 1185
246 GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262
1186 GCCACA.....ATTACTGTTTACTGTGGAGCTT 1214
262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuI 279
1215 TGGGCTCTGCAACAAGAGACTCCATAAAGGCATATGTGCTGGCTGCC 1264
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
1265 TGTATCCCTTTGGTGAATTACAGTACTGCTTATCAGAGAAAGCAAGCTT 1314
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
1315 CTCCCTCGAGCTGGAGAGACAGCCATCCAAATATACACTGTACAGGA 1364
312 rProGlnGluThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328

seq_name: /SIDS1/gcgdata/geneseq/NA2000.DAT:AAA91624

seq_documentation_block:

ID AAA91624 standard; DNA; 1642 BP.

AC AAA91624;

DT 12-JAN-2001 (first entry)

DE Caenorhabditis elegans cod-5 gene.

KW Caenorhabditis elegans; cod-5; daf-18; Insulin signalling pathway;

KW daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;

KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes; ds.

OS Caenorhabditis elegans.

XX Key Location/Qualifiers

XX CDS 44..1642

XX /*tag= a

XX /product= "Cod-5"

XX W0200033068-A1.

XX 08-JUN-2000.

XX 02-DEC-1999; 99WO-US28529.

XX 03-DEC-1998; 98US-0205658.

XX (GENO) GEN HOSPITAL CORP.

XX Ruvkun G, Ogg S;

XX WPI; 2000-423022/36.

XX P-PSDB; AAB13327.

XX Diagnosing and treating obesity and impaired glucose tolerance using

XX modulators of daf-18 expression and/or activity -

XX Disclosure; Fig 42; 402pp; English.

XX The present sequence is the cod-5 gene from *Caenorhabditis elegans*. Cod-5 is the aromatic amino acid hydroxylase that synthesizes serotonin from precursor L tryptophan. The cod-5 gene was knocked out to produce mutants completely lacking in serotonin. These mutants were found to have defects in metabolic control. A number of *C. elegans* proteins that have mammalian homologues acting in the insulin signalling pathway were also identified. The *C. elegans* age-1 gene encodes a homologue of the mammalian PI 3-kinase whilst daf-2 encodes a homologue of the mammalian insulin receptor. The *C. elegans* AKT kinase and PKB kinase act downstream of daf-2 and age-1, just as their mammalian homologues act downstream of insulin signalling. The *C. elegans* PTEN lipid phosphatase homologue, daf-18, has been found to act upstream of AKT in the pathway. This discovery has enabled mammalian PTEN action to be mapped to the insulin signalling pathway. Conserved DAF motifs can be used to design probes to identify mammalian DAF homologues and thus to identify individuals with a predisposition towards the development of glucose intolerance conditions, such as obesity and diabetes.

SQ Sequence 1642 BP; 479 A; 347 C; 351 G; 465 T; 0 other;

alignment_scores:

Quality: 206.50 Length: 278
Ratio: 1.424 Gaps: 8
Percent Similarity: 52.158 Percent Identity: 25.180

alignment_block:

US-09-438-185-1047 x AAA91624

Align seg 1/1 to: AAA91624 from: 1 to: 1642

104 ArgAsnLeuTrp.....TyrArgLeuLeuSerSerArgPheSerLe 117
||||| ||| ||||| |||
794 CGGAAACCTTGGGAATTATATAGAAATGAGA.....GAATF 834
117 uTrpLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
||||| ||| ||||| |||
835 GCACAAAGACGACGATGACAGCAGTTTCTTGTAACTTGGAGCTACTGG 884
134 lyLeuLeuSerAspPheLeuAspHis.....GlnAlaVal 145
885 AGAGACATTGTGATCTACTCGAAATAATATTCGCAACTAGAGATATC 934
146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerGl 162
||||| ||| ||||| |||
935 TCGAAGTTTGTAAAGCAAAACTGGATTCGCTGTCGCCACGTCGCCGG 984
162 yPheValAlaProHisGlnTyrLeuSerLeuGlnAspArgTyrPheP 179
||||| ||| ||||| |||
985 ATACTTATCAGCTCGTGATTTCTTGGCAGGCTTGCATATCGTGTCTCT 1034
179 rolLeAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
||||| ||| ||||| |||
1035 TCTGCACCTCAATACGTTCCGCATCATGCCGATCCATTTTACACTCCAGAA 1084
196 ProAspLeuIleHisAspLeuLeuGlyHisValProTrpLeuLeuHisPr 212
||||| ||| ||||| |||
1085 CCAGACACCGTTCCAGCGCTCATGGTCCATGCTCTATTCGCTGATCC 1134
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
||||| ||| ||||| |||
1135 AGATTTGCTCAGTTTCTCAAGAGATGGA.....T 1166
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
||||| ||| ||||| |||
1167 TAGCTTCTCTGGAGCATCAGAGGAGATTTGAAGAGCTTGCAACACTC 1216
246 GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262
||||| ||| ||||| |||
1217TACTTCTTTCCATTTGAATT 1236
262 rGlyLeuIle.....G 266
1237 TGGTCTCTGCTGATGCGCTGCCGATTCCTCCAGTAAAGAAATGGAT 1286
266 LuAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerPro 282
||||| ||| ||||| |||
1287 CAATCATGAAGATTTAAAGTATACGGAGCAGGACTTCTGAGCAGTGCT 1336
283 GlnGluLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGl 299
||||| ||| ||||| |||
1337 GCCGAGTTGCAACATCCGCTGAGGGTAGTCCACCATTTATCGTTTGA 1386
299 uLeuAspGlnIleArgLeuProPheAsnThrSerThrProGlnClut 316
||||| ||| ||||| |||
1387 TCCGATGCTGTTGTGACACAGATGCTCATCTACTACTTCCAGTCAG 1436
316 hrLeuPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLys 332
||||| ||| ||||| |||
1437 CGTATTCTTACTAGAAATTTGAAGAGGCCCGCAGCAG.....AAA 1477
333 LeuGluTrpMetLeuAspGlnGlyLeuLeuGluSerIleProLeuTyrAs 349
||||| ||| ||||| |||
1478 CTCAGAAATGTTCCACCAACCAACATGAACGTCCTTCATCTGTTGCTTACAA 1527
349 nGlnGluLysTyrLeuSerGlyPheGluValLeu 360
||||| ||| ||||| |||
1528 C.....CCATACACAGAAAGCGTCGAAGTTCTC 1555
seq_name: /SIDS1/gcgdata/geneseq/NA1993.DAT:AAQ41866

seq_documentation_block:
ID AAQ41866 standard; DNA; 1829 BP.
AC AAQ41866;
XX
XX 20-SEP-1993 (first entry)
XX
XX Human tyrosine hydroxylase coding sequence.
XX
XX Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;
KW neurological disorder; antibody; variant; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 20..1512
FT /tag= a
XX
XX J055212082-A.
XX
XX PD 18-MAY-1993.
XX
XX PF 13-MAR-1991; 91US-0669446.
XX
XX PR 13-MAR-1991; 91US-0669446.
XX
XX PA (UWNY) UNIV NEW YORK STATE.
XX
XX PI Filer D, Friedhoff AJ, Goldstein M, Wu J;
XX
XX DR WPI; 1993-175456/21.
XX
XX DR P-PSDB; AAR36740.
XX
XX Genetically modified tyrosine hydroxylase having increased
PT activity - used for treating neurological disorders e.g.
PT Parkinson's and Alzheimer's diseases and affective disorders
XX
XX PS Disclosure; Fig 6; 20pp; English.
XX
XX This sequence encodes the human tyrosine hydroxylase protein. This
CC sequence was used in the construction of a variant tyrosine hydroxylase
CC which contains at least one amino acid substitution in the N-terminal
CC 55 amino acids. The substitution corresponds to an amino acid selected
CC from Ser8, Arg31, Arg37, Arg38, Gln39, Ser40, Leu42, Ile42, Glu43,
CC Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant
CC protein is at least 3-fold greater than that for the wild type
CC protein. Cells transfected with this DNA may be used for treating
CC neurological disorders associated with a deficiency in tyrosine
CC hydroxylase or dopamine. The tyrosine hydroxylase protein may be
CC used to generate antibodies specific for the variant tyrosine
CC hydroxylases to monitor the enzyme during a treatment regimen.
XX
XX Sequence 1829 BP; 309 A; 636 C; 559 G; 325 T; 0 other;

alignment_scores:
Quality: 198.00 Length: 219
Ratio: 1.623 Gaps: 4
Percent Similarity: 55.708 Percent Identity: 26.941
alignment_block:
US-09-438-185-1047 x'AAQ41866 ..
Align seg 1/1 to: AAQ41866 from: 1 to: 1829
128 ASPTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis..... 142
766 GAGCACCTGGAGCGCTTCCTTTGCTGGAGCGCTTCAGCGGCTACCGGGA 815
143GlnAlaValIleLysPheGluLeuGluT 153
816 AGACATATATCCCCAGCTGGAGGAGGCTCCCGCTTCCTGAGGAGCGCA 865

153 hrHisPheSerTyrTyrProValSerGlyPheValAlaProHisGlnTyr 169
866 GGGGCTTCCAGCTGCGGGCTGTGGCGGCTGCTGTCCCGCGGACTTC 915
170 LeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerValMetArgTh 186
916 CTGGCAGCCTGGCCTTCGCGGTGTTCCAGTGCACCCAGTATATCCGCCA 965
186 rleuAspLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeuL 203
966 CGCGTCCVCGCCCATGCACTCCCTGAGCGGACTGCTGCCACGAGCTGC 1015
203 euGlyHisValProTyrPheLeuHisProSerPheSerGluPheHeile 219
1016 TGGGACAGTCCCATGCTGCGCGACCGCACTTCGCGCAGTTCGCGAG 1065
220 AsnMetGlyArgLeuPheThrLysValIleGluLysValGlnAlaLeuPr 236
1066 GACATTGGC.....CTGGCGTCCCTGGGGCGCTCGGA 1097
236 oSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIleAlaIleV 253
1098 TGAGGAATTCGAGAGCTGTCCACGCTGCA..... 1128
253 alArgCysPheThrPheThrValcIuSerGlyLeuIleGluAsnHisGlu 269
1129TGTTTCCAGTGGAGTTCGGGCTGTGTAAAGCAGAACCGG 1167
270 GlyArgLysAlaTyrGlyAlaValLeuIleSerProGlnGlnLeuG1 286
1168 GAGGTGAAGCCCTATGTGTGGCGGCTGCTCTCTCCTACGGGGAGCTCCT 1217
286 yHisAlaPheIleAspAsnValArgValLeuProLeuGluLeuAspGlnI 303
1218 GCACCTGCCTGTCTGAGGAGCTGAGATTTCGGGCTTCGACCTGAGGCTG 1267
303 leIleArgLeuProPheAsnThrSerThrProGlnGlnThrLeuPheSer 319
1268 CGGCGGTGCAGCCCTACCAAGACAGCAGCTACCACTGCTACTTCGTG 1317
320 IleArgHisPhe.....AspGluLeuValGluLeuThrSerLy 332
1318 TCTGAGAGCTTCAGTGACGCGCAGGACAGCTCAGGAGCTATCCCTCAGG 1367
332 sleuGlu 334
1368 CAIICAG 1374
seq_name: /SDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAA91625
seq_documentation_block:
ID AAA91625 standard; cDNA; 817 BP.
XX
XX AC AAA91625;
XX
XX 12-JAN-2001 (first entry)
XX
XX Caenorhabditis elegans cod-5 knockout cDNA.
XX
XX Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;
KW daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;
KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes; ss.
XX
XX OS Caenorhabditis elegans.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT CDS 40..177
FT /tag= a
FT /product= "Cod-5 frameshift mutant"
XX
XX PN W0200033068-A1.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:51:51 ; Search time 31 Seconds
(without alignments)
400.016 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889
Sequence: 1 VHYCERTLPKYLKALKL.....ESIPLYNQEKYLSGFVQLCQ 362

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235.5	12.5	262	1 PH4H_PSEAE	P43334 pseudomonas
2	231	12.2	445	1 TR5H_CHICK	P70080 gallus gall
3	230.5	12.2	444	1 TR5H_HUMAN	P17752 homo sapien
4	229.5	12.1	491	1 TY3H_PHASP	P1982 phasianidae
5	227.5	12.0	444	1 TR5H_RAT	P09810 rattus norv
6	226.5	12.0	447	1 TR5H_MOUSE	P17532 mus musculu
7	225.5	11.9	444	1 TR5H_RABIT	P17290 oryctolagus
8	225.5	11.9	488	1 TY3H_ANGAN	O42091 anguilla an
9	225.5	11.9	498	1 TY3H_MOUSE	P24529 mus musculu
10	223.5	11.8	481	1 TR5H_XENLA	Q92142 xenopus lae
11	223	11.8	498	1 TY3H_RAT	P04177 rattus norv
12	221.5	11.7	490	1 TY3H_BOVIN	P17289 bos taurus
13	219	11.6	528	1 TY3H_HUMAN	P07101 homo sapien
14	219	11.6	508	1 TY3H_DROME	P18459 drosophila
15	216.5	11.5	465	1 TY3H_SCHMA	O17446 schistosoma
16	214	11.3	453	1 PH4H_RAT	P04176 rattus norv
17	213.5	11.3	452	1 PH4H_HUMAN	P00439 homo sapien
18	212.5	11.2	453	1 PH4H_DROME	P17276 drosophila
19	194	10.3	452	1 PH4H_CHRVO	P30967 chromobacte
20	190	10.1	281	1 PH4H_CABEL	P30986 caenorhabdi
21	178.5	9.4	404	1 TY3H_CABEL	P30925 caenorhabdi
22	175.5	9.3	457	1 NINL_DROME	P10676 drosophila
23	166.5	5.6	1501	1 RYSL_MABVP	P35262 marburg vir
24	104.5	5.5	2331	1 RYSL_SCHPO	Q10144 schizosacch
25	102	5.4	563	1 YASL_SCHPO	Q36161 saccharomyc
26	100.5	5.3	1157	1 N133_YEAST	Q05318 ebola virus
27	100.5	5.3	2212	1 RRPL_EBOZM	P47321 mycoplasma
28	98.5	5.2	1024	1 Y075_MYCGE	O27712 panulirus a
29	97	5.1	492	1 CPL1_PANAR	Q92767 chlamydia p
30	97	5.1	1050	1 EX5B_CHLPN	Q82806 chlamydia p
31	96.5	5.1	512	1 G6PD_CHLPN	Q09812 schizosacch
32	96	5.1	529	1 YABD_SCHPO	P52362 human herpe
33	96	5.1	2059	1 TEGU_HSV7J	

34	95.5	5.1	1398	1 TOP2_PLAPK	P41001 plasmodium
35	95	5.0	1427	1 SRB8_YEAST	P25648 saccharomyc
36	95	5.0	1475	1 TRA2_CABEL	P34709 caenorhabdi
37	93.5	4.9	650	1 Y096_MYCGE	P47342 mycoplasma
38	92.5	4.9	466	1 Y065_MYCGE	P47311 mycoplasma
39	92	4.9	447	1 FLP_KLULA	P13783 kluyveromyc
40	92	4.9	526	1 MATK_ALATH	P56784 arabidopsis
41	92	4.9	544	1 MATK_MAIZE	P48190 zea mays (m
42	92	4.9	4563	1 APB_HUMAN	P04114 homo sapien
43	91.5	4.8	293	1 SPEE_METJA	Q57761 measles vir
44	91	4.8	2183	1 RRPL_MEASA	P35975 measles vir
45	91	4.8	2183	1 RRPL_MEASE	P12576 measles vir

ALIGNMENTS

RESULT 1
PH4H_PSEAE
ID PH4H_PSEAE STANDARD; PRT: 262 AA.
AC P43334;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PHENYLALANINE-4-HYDROXYLASE (EC 1.14.16.1) (PAH) (PHE-4-MONOOXYGENASE).
DE PHA OR PA0872.
GN Pseudomonas aeruginosa.
OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 15692 / PA01;
RC MEDLINE=94151331; PubMed=8108417;
RA Zhao G., Xia T., Song J., Roy R.A.;
RT "Pseudomonas aeruginosa possesses homologues of mammalian phenylalanine hydroxylase and 4 alpha-carbinolamine dehydratase/DCoH as part of a three-component gene cluster."
RL Proc. Natl. Acad. Sci. U.S.A. 91:1366-1370(1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=PA01;
RC MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Coulter S.N., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Goulet S., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- CATALYTIC ACTIVITY: L-PHENYLALANINE + TETRAHYDROBIOTERIN + O(2) -> L-TYROSINE + DIHYDROBIOTERIN + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: CATABOLISM OF PHENYLALANINE; FIRST, RATE-LIMITING STEP.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M88627; AAA25936.1;
CC EMBL: A800452; BAG04261.1;
CC HSSP: P04177; 1TQH
CC InterPro: IPR001273; --

DR Pfam: PF00351; biotin.H; 1.
DR PROSITE: PS00367; BIOTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron.
FT METAL 121 121 IRON (BY SIMILARITY).
FT METAL 126 126 IRON (BY SIMILARITY).
FT CONFLICT 135 135 F -> L (IN REF. 1).
SQ SEQUENCE 262 AA; 30322 MW; A565839C5961A45 CRC64;

Query Match 12.5%; Score 235.5; DB 1; Length 262;
Best Local Similarity 24.3%; Pred. No. 9e-12;
Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;

QY 93 DMAYVSTFFNNRWLLSRSLKSYCPREFLDYLEAFGLLDFDLDHQAIVKFELE 152
DB 13 DNGFHYETHEWNTLTQLKLVIEGRACQYLGIEQLG-----LPHRIPOLEIN 67
QY 153 -----THFSYYPVSGFVAPHOYLSLLQDRFPFIASVMRTLDKNFSLTPDLIHDLLGHV 206
DB 68 RVLQATQGWVRVAPALIPFQTFEELLASQFPVATFIRPELDYLGQEPDIFHEIFGHC 127
QY 207 PWWLHPSEFFNNGRLFTKVIKQVQALPSKKORIOTLQSNLIAIVRCFWFTVESGLIE 266
DB 128 PLTNPWFAEFTYTGKLGKA-----SKER-----VFLARLYNMTIEFGLVE 171
QY 267 NHEGRKAYGAVLLISPOELGHAFID---NVRVLPLELDQIIRLPENTSTPOETLFSIRHF 323
DB 172 TDQGRYVGGILSPKRTVYSLSDEPLHQAENPLE---AMRTPYRIDILQPLVFLPDL 228
QY 324 DELVELTSK 332
DB 229 KRLFLAQE 237

RESULT 2
ID TR5L CHICK STANDARD; PRT; 445 AA.
AC P70860;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC Gallus.
CC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-Pineal gland;
RX MEDLINE=97072811; PubMed=8915576;
RA Florez J.C., Seidenman K.J., Barrett R.K., Sangoram A.M.,
RA Takahashi J.S.;
RT "Molecular cloning of chick pineal tryptophan hydroxylase and
RT circadian oscillation of its mRNA levels";
RL Brain Res. Mol. Brain Res. 42:25-30(1996).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
CC EMBL: U26428; AAC60036.1;
DR HSP: P04177; ITOH.
DR InterPro: IPR001273;
DR InterPro: IPR002912;
DR Pfam: PF01842; ACT.1.
DR Pfam: PF00351; biotin.H; 1.
DR PRINTS: PR00372; FYWHYDRXLA.
DR PROSITE: PS00367; BIOTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation. 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT METAL 273 273 IRON (BY SIMILARITY).
FT METAL 278 278 IRON (BY SIMILARITY).
FT METAL 318 318 IRON (BY SIMILARITY).
SQ SEQUENCE 445 AA; 51139 MW; FF0041D7C4B159F6 CRC64;

Query Match 12.2%; Score 231; DB 1; Length 445;
Best Local Similarity 28.4%; Pred. No. 4.1e-11;
Matches 65; Conservative 41; Mismatches 93; Indels 30; Gaps 4;

QY 108 YRLSSRFSLKSYCPREFLDYLEAFGLLSDFLDH-----QAVIKFELETHESYYP 159
DB 179 YRELN-----KLYTHACREYLNKLNPLLTLYCYGYREDNIPQLEDVSRFLKERTGTIRP 232
QY 160 VSGFVAPHOYLSLLQDRFPFIASVMRTLDKNFSLTPDLIHDLLGHVPLHPSFSEFF 219
DB 233 VAGYLSPRDFLAGLAFRVFHCQYVRRHSSDPLYTPEDTCHELLGHVPLLAEPFAQFSQ 292
QY 220 NMGRLETKVIEKQVQALPSKKORIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLI 279
DB 293 EIG-----LASLGASDEAVOKLAT-----CYFTVEFGLCQEGQLRVYGNLL 336
QY 280 SSPQELGHAFIDNVRVLPLELDQIIRLPENTSTPOETLFSIRHDELVE 328
DB 337 SSISELAHSLSGAKVPPDPKVTCKQECCLITTFQEVYFVSSEFEAKE 385

RESULT 3
ID TR5H HUMAN STANDARD; PRT; 444 AA.
AC P17752; Q16736;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH OR TRPH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Carcinoma;
RX MEDLINE=90332431; PubMed=2377472;
RA Boulard S., Darmon M.C., Ganem Y., Launay J.M., Mallet J.;
RT "Complete coding sequence of human tryptophan hydroxylase";
RL Nucleic Acids Res. 18:4257-4257(1990).
CC [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=95077422; PubMed=7986090;
RT Tipper J.P., Citron B.A., Ribeiro P., Kaufman S.;
RT "Cloning and expression of rabbit and human brain tryptophan
RT hydroxylase cDNA in Escherichia coli";
RL Arch. Biochem. Biophys. 315:445-453(1994).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST

CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEL GLAND.
CC -1- SUBUNIT: MOLIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52836; CAA37018.1; -;
DR EMBL; L29306; AAA67050.1; -;
DR PIR; S10489; S10489.
DR HSP; P04177; ITOH.
DR MIM; 191060; -;
DR InterPro: IPR001273; -;
DR InterPro: IPR002912; -;
DR Pfam: PF01842; ACT. 1.
DR PRINTS: PR00351; bioprotein.H; 1.
DR PRINTS: PR00372; FWHYDEXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
FT METAL 319 319 IRON (BY SIMILARITY).
FT CONFLICT 19 19 T -> S (IN REF. 1).
FT CONFLICT 68 68 T -> I (IN REF. 1).
FT CONFLICT 90 91 TP -> NL (IN REF. 1).
FT CONFLICT 97 97 M -> L (IN REF. 1).
FT CONFLICT 100 100 E -> D (IN REF. 1).
FT CONFLICT 104 104 S -> T (IN REF. 1).
FT CONFLICT 151 151 L -> S (IN REF. 2).
FT CONFLICT 154 154 N -> S (IN REF. 2).
FT CONFLICT 157 157 H -> Y (IN REF. 2).
FT CONFLICT 179 179 R -> Q (IN REF. 1).
FT CONFLICT 207 207 R -> Q (IN REF. 2).
FT CONFLICT 217 217 V -> I (IN REF. 2).
FT CONFLICT 344 344 A -> V (IN REF. 2).
FT CONFLICT 414 414 T -> A (IN REF. 2).
FT CONFLICT 419 419 S -> N (IN REF. 2).
FT CONFLICT 425 425 R -> Q (IN REF. 1).
FT CONFLICT 436 436 A -> G (IN REF. 2).
SQ SEQUENCE 444 AA; 51032 MW; 86C39884F72E120A CRC64;
Query Match 12.2%; Score 230.5; DB 1; Length 444;
Best Local Similarity 27.8%; Pred. No. 4.5e-11;
Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;
QY 107 WYLLSSRSLKWSYCPREFLDYLEAFGLSLDFLDH-----QAVIKFFETHFSY 158
Db 174 WGTVFRELKLYPTHACR---EYLKNEPLLSKYGYREDNIPQLEDVSNFLKERTGFSIR 230
QY 159 PVSGFVAPHOYLSLDQRYFPPIASVMRTLDKDNFSLPDLIHLGLHVPWLLHPSFSFF 218
Db 231 PVAGYLSRDLFSLGLAFVHCTQYVRRHSDPFTPEPDTCHELLGHVPLLAEPFAQFS 290
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSLNLAIVRCFWFTVSGLIENHEGRKAYGAVL 278
Db 291 QEIG-----LASIGASEAVQKAT-----CYFTVEFGKODGQLRVFAGL 334
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPFNFTSTPQETLSIRHFDE 325
Db 335 LSSISELKHLSGHAKVFPDKPTCKOECLITTFQDVYVFSSEFDEAKE 384

RESULT 4
TY3H_PHASP

ID TY3H_PHASP STANDARD; PRT; 491 AA.
AC P11982; 1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 35, Last annotation update)
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).
GN TH.
OS Phasianidae sp. (Quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae.
OX NCBI_TaxID=9006;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RX MEDLINE=88089590; PubMed=2447231;
RA Fauquet M., Grima B., Lamouroux A., Mallet J.;
RT "Cloning of quail tyrosine hydroxylase: amino acid homology with
other hydroxylases discloses functional domains.";
RL J. Neurochem. 50:142-148(1988).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24778; AAA49514.1; -;
DR PIR; A28582; A28582.
DR HSP; P04177; ITOH.
DR InterPro: IPR001273; -;
DR Pfam: PF00351; bioprotein.H; 1.
DR PRINTS: PR00372; FWHYDEXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis; Phosphorylation.
FT MOD_RES 40 40 PHOSPHORYLATION (BY CAPK)
FT MOD_RES 40 40 (BY SIMILARITY).
FT DOMAIN 51 55 POLY-ALA.
FT METAL 324 324 IRON (BY SIMILARITY).
FT METAL 329 329 IRON (BY SIMILARITY).
FT METAL 369 369 IRON (BY SIMILARITY).
SQ SEQUENCE 491 AA; 56066 MW; AFB363220F70C0A0 CRC64;
Query Match 12.1%; Score 229.5; DB 1; Length 491;
Best Local Similarity 29.1%; Pred. No. 6.1e-11;
Matches 66; Conservative 34; Mismatches 100; Indels 27; Gaps 4;
QY 107 WYLLSSRSLKWSYCPREFLDYLEAFGLSLDFLDH-----QAVIKFFETHFSY 158
Db 256 WKEYVSTLSLYPTHACK---EYLEAFNLKFCGYNNNTIPQLEEVSRFLKERTGFQRL 282
QY 159 PVSGFVAPHOYLSLDQRYFPPIASVMRTLDKDNFSLPDLIHLGLHVPWLLHPSFSFF 218
Db 283 PVRLLSARDFLASLAFRVQCTQYIRHASSPMHSPEDCCHELLGHVPLADKTFQFS 342
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSLNLAIVRCFWFTVSGLIENHEGRKAYGAVL 278
Db 343 QDIG-----LASIGADEIEKATL-----YFTVEFGICRQNGIVKAGACL 386
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPFNFTSTPQETLSIRHFDE 325

Db 387 LSSYGEIHLISLDEVRDFDPAADAAVOPCQDPQPVYFVSEFSD 433

```
RESULT 5
TRSH_RAT TRSH_MOUSE STANDARD; PRT; 444 AA.
AC P09810;
DT 01-WAR-1989 (Rel. 10, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN-WISTAR; TISSUE=Pineal gland;
RX MEDLINE=88244702; PubMed=3379411;
RA Darmon M.C., Guilbert B., Levieil V., Ehret M., Maitre M., Mallet J.;
RT "Sequence of two mRNAs encoding active rat tryptophan hydroxylase.";
RL J. Neurochem. 51:312-316(1988).
RN [2]
SEQUENCE FROM N.A.
RA MEDLINE=91245924; PubMed=1645430;
RX Kim K.S., Wessel T.C., Stone D.M., Carver C.H., Joh T.H., Park D.H.;
RT "Molecular cloning and characterization of cDNA encoding tryptophan
hydroxylase from rat central serotonergic neurons.";
RL Brain Res. Mol. Brain Res. 9:277-283(1991).
RN [3]
SEQUENCE OF 167-261 FROM N.A.
RA MEDLINE=87005247; PubMed=2875901;
RX Darmon M.C., Grima B., Cash C.D., Maitre M., Mallet J.;
RT "Isolation of a rat pineal gland cDNA clone homologous to tyrosine
and phenylalanine hydroxylases.";
RL FEBS Lett. 206:43-46(1986).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; M28000; AAA2262.1; -
DR EMBL; X53501; CAA37579.1; -
DR PIR; J10034; WHRTW.
DR PIR; A24367; A24367.
DR PIR; A60034; A60034.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; -
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT METAL 272 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT METAL 277 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
```

FT METAL 317 317 IRON (BY SIMILARITY).
SQ SEQUENCE 444 AA; 51068 MW; C3CF5245727CC825 CRC64;

```
Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 7.7e-11;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;
QY 107 WYRLSSRFSLKSYCPFFLDYLEAFGLSLDFLDH-----QAVIKPFELETHFSY 158
DB 174 WGTFRRLNKLYTHACR--EYLRNPLLSKYCGYREDNVPOLEDVSNELKERTGFSIR 230
QY 159 PVSGFVAPHOYLSLODRYPPTASVMTLDKDNFSLTDLHDLGHVPWLLHPSEFF 218
DB 231 PVAGYLSRPFDSGLAFRVFHCYQVYRHSDDPLYTPEDTCHELLGHVPLLAEPFSAQS 290
QY 219 INMGRLFTKVIKRVQALPSKKQRIQTLQSLNLAIVRCFWFTVSEGLIENHEGRKAYGAVL 278
DB 291 QEIG-----LASLGASEETVQKLAT-----CYFTTFEGLCKQDQQRVFGAGL 334
QY 279 ISSPQELGHAFIDNVRYLPLELDQIIRLPFNFTSTPQETLFSIRHFDLVE 328
DB 335 LSSISELRLHALSGHAKVFPDPKACKQELITSFQDVYFVSEFDEAKE 384
```

```
RESULT 6
TRSH_MOUSE STANDARD; PRT; 447 AA.
AC P17532;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=90243261; PubMed=2110547;
RX Stoll J., Kozak C.A., Goldman D.;
RT "Characterization and chromosomal mapping of a cDNA encoding
tryptophan hydroxylase from a mouse mastocytoma cell line.";
RL Genomics 7:88-96(1990).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; J04758; AAA63401.1; -
DR PIR; A34582; A34582.
DR HSSP; P04177; ITOH.
DR MGD; MGI:98796; Tph.
DR InterPro; IPR001273; -
DR InterPro; IPR002912; -
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
```

KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 61 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT METAL 275 IRON (BY SIMILARITY).
FT METAL 280 IRON (BY SIMILARITY).
FT METAL 320 IRON (BY SIMILARITY).
SQ SEQUENCE 447 AA; 51343 MW; 16C839F22A138BCA CRC64;

Query Match 12.0%; Score 226.5; DB 1; Length 447;
Best Local Similarity 27.0%; Pred. No. 9.3e-11;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKSYCPREFLDYLEAFGLSDFLDH-----QAVIKFFELTHFSY 158
Db 177 WGTIFRELKLYPTHACR---EYLKPLLSKYCYREDNIPQLEDVSNFLKERTGSIR 233
QY 159 PVSGFVAPHQYLSLQDRYFPIASVMTLQKDNFSLTPDLIHLLGHVPLWLLHPSFEFF 218
Db 234 PVAGYLSRDFLSGLAFRVFCHTQYVHRSSDPLTPEDTCHELLGHVPLLAEPFSAQFS 293
219 INNGRLFTVKIEKVALPSKKORIOTLOSNIAlVRCFWFTVESGLIENHGRKAYGAVL 278
294 QEIG-----LASLGASETVOKIAT-----CYFFTFEFLCKODGQLRVFGAGL 337
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328
Db 338 LSSISELKHVLGHAKVPDPKITYKOECLITTFQDVYFVSFEDAKE 384

RESULT 7
TRSH_RABIT STANDARD; PRT; 444 AA.
AC P17290; Q29523;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87289638; PubMed=3475690;
RA Grenett H.E., Ledley F.D., Reed L.L., Woo S.L.C.;
FT "Full-length cDNA for rabbit tryptophan hydroxylase: functional
domains and evolution of aromatic amino acid hydroxylases.";
Proc. Natl. Acad. Sci. U.S.A. 84:5530-5534(1987).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95077422; PubMed=7986090;
RA Tipper J.P., Citron B.A., Ribeiro P., Kaufman S.;
FT "Cloning and expression of rabbit and human brain tryptophan
hydroxylase cDNA in Escherichia coli.";
Arch. Biochem. Biophys. 315:445-453(1994).
RL Arch. Biochem. Biophys. 315:445-453(1994).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOTRIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M17250; AAA31487.1; -;
DR EMBL; L29305; AAA67051.1; -;
DR PIR; A32699; A32699.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; -;
DR InterPro; IPR002912; -;
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; biotinH; 1.
DR PRINTS; PR00372; FYNHYDRXLASE.
DR PROSITE; PS00367; BIOTRIN_HYDROXYL; 1.
DR Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
KW MOD_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
FT CONFLICT 102 102 M -> L (IN REF. 1).
FT CONFLICT 151 151 L -> S (IN REF. 2).
FT CONFLICT 151 151 KY -> ND (IN REF. 1).
FT CONFLICT 202 203 R -> Q (IN REF. 2).
FT CONFLICT 207 207 T -> K (IN REF. 1).
FT CONFLICT 390 390
SQ SEQUENCE 444 AA; 51118 MW; BF182451B28ECD80 CRC64;

Query Match 11.9%; Score 225.5; DB 1; Length 444;
Best Local Similarity 27.0%; Pred. No. 1.1e-10;
Matches 62; Conservative 44; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKSYCPREFLDYLEAFGLSDFLDH-----QAVIKFFELTHFSY 158
Db 174 WGTIFRELKLYPTHACR---EYLKPLLSKYCYREDNIPQLEDVSNFLKERTGSIR 230
QY 159 PVSGFVAPHQYLSLQDRYFPIASVMTLQKDNFSLTPDLIHLLGHVPLWLLHPSFEFF 218
Db 231 PVAGYLSRDFLSGLAFRVFCHTQYVHRSSDPLTPEDTCHELLGHVPLLAEPFSAQFS 290
QY 219 INNGRLFTVKIEKVALPSKKORIOTLOSNIAlVRCFWFTVESGLIENHGRKAYGAVL 278
Db 291 QEIG-----LASLGASETVOKIAT-----CYFFTFEFLCKODGQLRVFGAGL 334
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328
Db 335 LSSISELKHVLGHAKVPDPKITYKOECLITTFQDVYFVSFEDAKE 384

RESULT 8
TY3H_ANGAN STANDARD; PRT; 488 AA.
ID TY3H_ANGAN
AC O42091;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).
GN TH.
OS Anguilla anguilla (European freshwater eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC Anguillidae; Anguilla.
OX NCBI_TaxID=7936;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98344760; PubMed=9681435;
RA Boularand S., Biguet N.F., Vidal B., Veron M., Mallet J.,
RA Vincent J.D., Dufour S., Vernier P.;
FT "Tyrosine hydroxylase in the European eel (*Anguilla anguilla*): cDNA
cloning, brain distribution, and phylogenetic analysis.";
J. Neurochem. 71:460-470(1998).
RL J. Neurochem. 71:460-470(1998).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =

```
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY (BY SIMILARITY).
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ000731; CAA04264.1; -.
CC HSSP: P04177; ITOH.
CC InterPro: IPR001273; -.
CC Pfam: PF00351; biopterin_H; 1.
CC PRINTS: PR00367; FWHYDRXKLASE.
CC PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
CC Neurotransmitter biosynthesis; Phosphorylation.
CC MOD_RES 38 38 PHOSPHORYLATION (BY CAPK) (BY
CC SIMILARITY).
CC METAL 321 321 IRON (BY SIMILARITY).
CC METAL 326 326 IRON (BY SIMILARITY).
CC METAL 366 366 IRON (BY SIMILARITY).
CC SEQUENCE 488 AA; 55490 MW; 573553BA39EBD448 CRC64;
CC -----
Query Match 11.9%; Score 225.5; DB 1; Length 488;
Best Local Similarity 28.1%; Pred. No. 1.3e-10;
Matches 65; Conservative 38; Mismatches 99; Indels 29; Gaps 5;
QY 107 WYRLSSRFLSKSY-CPRFLDYLLEAFGLLSDFLDH-----QAVIKFELETHFSY 157
Db 223 WREYVSTLDLYTHACS---EHLEAFLLERHCYSPNSIPOLEDVSHLKERTGFQL 278
QY 158 YVSGFVAPHOYLSLLQDRYFPFIASVMRTLDKDNFSLTDLIHLGHVPMWLLHPSFSEF 217
Db 279 RPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCHELLGHVPMADRTAQAQ 338
QY 218 FTMGRLETKVIEKQVQALPSKKQRIOTQLOSLNLIIVRCFWFTVESGLIENHEGRKAYGAV 277
339 SQNIG-----LASLGASEEDIEKLSL-----LWFTVEFGKQGDGVKAYGAG 382
QY 278 LSSPQELGHAFIDNVRLPLELDQIIRLPNTSTPQETFLFSIRHDELVE 328
Db 383 LSSYGLVHLSDEPERREFDPEAAAEYQDQNYQSVYFVSESFTDAKE 433
RESULT 9
TY3H_MOUSE STANDARD; PRT: 498 AA.
AC P24529;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).
GN TH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91248263; PubMed=1674869;
RA Chikawa S., Sasaki A., Nagatsu T.;
RT Primary structure of mouse tyrosine hydroxylase deduced from its
RT cDNA.;
RL Biochem. Biophys. Res. Commun. 176:1610-1616(1991).
```

```
RN SEQUENCE OF 1-30 FROM N.A.
RP STRAIN=BALB/C;
RA Morgan W.W., Bermudez J., Sharp Z.D.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + DIHYDROPTERIDINE + O(2) -
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M69200; AAA40434.1; -.
CC EMBL: X53503; CAA37580.1; -.
CC PIR: JN0068; JN0068.
CC HSSP: P04177; ITOH.
CC MGD: MGI:98735; TH.
CC InterPro: IPR001273; -.
CC Pfam: PF00351; biopterin_H; 1.
CC PRINTS: PR00372; FWHYDRXKLASE.
CC PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
CC Neurotransmitter biosynthesis; Phosphorylation.
CC MOD_RES 19 19 PHOSPHORYLATION (BY CAM-KINASE II)
CC (BY SIMILARITY).
CC MOD_RES 31 31 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 40 40 PHOSPHORYLATION (BY CAPK)
CC (BY SIMILARITY).
CC DOMAIN 51 59 POLY-ALA.
CC METAL 331 331 IRON (BY SIMILARITY).
CC METAL 336 336 IRON (BY SIMILARITY).
CC METAL 376 376 IRON (BY SIMILARITY).
CC SEQUENCE 498 AA; 55992 MW; 62790179664F6DC6 CRC64;
CC -----
Query Match 11.9%; Score 225.5; DB 1; Length 498;
Best Local Similarity 27.5%; Pred. No. 1.3e-10;
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;
QY 107 WYRLSSRFLSKSY-CPRFLDYLLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
Db 233 WKEYATLKLGYATHACR---EHLEAFQLLERYCYGREDSTPOLEDVSHLKERTGFQLR 289
QY 159 YVSGFVAPHOYLSLLQDRYFPFIASVMRTLDKDNFSLTDLIHLGHVPMWLLHPSFSEFF 218
Db 290 YVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCHELLGHVPMADRTAQAQS 349
QY 219 INMGRLETKVIEKQVQALPSKKQRIOTQLOSLNLIIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 350 QDIG-----LASLGASEEDIEKLSL-----VWFTVEFGKQNGELKAYGAGL 393
QY 279 ISSPQELGHAFIDNVRLPLELDQIIRLPNTSTPQETFLFSIRHDELVE 334
Db 394 LSSYGLVHLSDEPERREFDPEAAAEYQDQNYQSVYFVSESFTDAKLNAYASRIQ 453
RESULT 10
TR5H_XENLA
ID TR5H_XENLA STANDARD; PRT: 481 AA.
AC Q92142;
DT 15-JUL-1998 (Rel. 36, Created)
```


AC P07101; Q15588; Q15589; Q15585;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE TYROSINE 3-MONOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).
GN TH OR IVH
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (TYPES 1 TO 4).
RX MEDLINE=87173064; PubMed=2862428;
RA Grima B., Lamouroux A., Boni C., Julien J.-F., Javoy-Agid F.,
RA Mallet J.;
RA "A single human gene encoding multiple tyrosine hydroxylases with
RT different predicted functional characteristics.";
RL Nature 326:707-711(1987).
RN [2]
RP SEQUENCE FROM N.A. (TYPE 3).
RX MEDLINE=87316931; PubMed=2888085;
RA Kobayashi K., Kaneda N., Ichinose H., Kishi F., Nakazawa A.,
RA Kurosawa Y., Fujita K., Nagatsu T.;
RA "Isolation of a novel cDNA clone for human tyrosine hydroxylase:
RT alternative RNA splicing produces four kinds of mRNA from a single
RL gene.";
RN Biochem. Biophys. Res. Commun. 146:971-975(1987).
RN [3]
RP SEQUENCE OF 1-94 FROM N.A. (TYPE 3).
RX MEDLINE=87298614; PubMed=2887169;
RA Kaneda N., Kobayashi K., Ichinose H., Kishi F., Nakazawa A.,
RA Kurosawa Y., Fujita K., Nagatsu T.;
RA "Isolation of a novel cDNA clone for human tyrosine hydroxylase:
RT alternative RNA splicing produces four kinds of mRNA from a single
RL gene.";
RN Biochem. Biophys. Res. Commun. 146:971-975(1987).
RN [4]
RP SEQUENCE FROM N.A. (TYPES 1 TO 4).
RX MEDLINE=89008200; PubMed=2902075;
RA Kobayashi K., Kaneda N., Ichinose H., Kishi F., Nakazawa A.,
RA Kurosawa Y., Fujita K., Nagatsu T.;
RA "Structure of the human tyrosine hydroxylase gene: alternative
RT splicing from a single gene accounts for generation of four mRNA
RL types.";
RN J. Biochem. 103:907-912(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88117543; PubMed=2892893;
RA le Bourdellies B., Boularand S., Boni C., Horellou P., Dumas S.,
RA Grima B., Mallet J.;
RA "Analysis of the 5' region of the human tyrosine hydroxylase gene:
RT combinatorial patterns of exon splicing generate multiple regulated
RL tyrosine hydroxylase isoforms.";
RN J. Neurochem. 50:988-991(1988).
RN [6]
RP SEQUENCE OF 1-30 FROM N.A.
RX MEDLINE=88213428; PubMed=2896667;
RA Gines E.I., Rehavi M., Martin B.M., Weller M., O'Malley K.L.,
RA Lamarca M.E., McAllister C.G., Paul S.M.;
RA "Expression of human tyrosine hydroxylase cDNA in vertebrate cells
RT using a baculovirus vector.";
RL J. Biol. Chem. 263:7406-7410(1988).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: RATE-LIMITING STEP IN CAPECHOLAMINE BIOSYNTHESIS.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y00414; CAA68472.1; ALT_SEQ.
DR EMBL: M17585; AAA611179.1; -
DR EMBL: X05290; CAA28908.1; -
DR EMBL: M24791; AAA611173.1; -
DR EMBL: M24787; AAA611173.1; JOINED.
DR EMBL: M24789; AAA611173.1; JOINED.
DR EMBL: M24791; AAA611170.1; -
DR EMBL: M24787; AAA611170.1; JOINED.
DR EMBL: M20911; AAA611167.1; -
DR PIR: A27791; WHHUY1.
DR PIR: B27791; WHHUY2.
DR PIR: C27791; WHHUY3.
DR PIR: A26825; WHHUY4.
DR PIR: JE0012; JE0012.
DR PIR: JE0013; JE0013.
DR PIR: JE0014; JE0014.
DR HSSP: P04177; ITOH.
DR MIM: 191290; -
DR InterPro: IPR001273; -
DR Pfam: PF00351; biotin.H.1.
DR PRINTS: PR00372; FWHVDRXIAE.
DR PROSITE: PS00367; BIOTPTXIN_HYDROXYL.1.
KW Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Alternative splicing; Neurotransmitter biosynthesis; Phosphorylation.
FT MOD_RES 71 71
FT DOMAIN 85 90
FT METAL 361 361
FT METAL 366 366
FT METAL 406 406
FT METAL 406 406
FT VARSPIC 1 33
FT VARSPIC 34 34
FT VARSPIC 31 61
FT VARSPIC 35 61
SQ SEQUENCE 528 AA; 58524 MW; B3DC349955A21074 CRC64;

Query Match 11.6%; Score 220; DB 1; Length 528;
Best Local Similarity 26.7%; Pred. No. 3.8e-10;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

QY 112 SSFSLWKS-----YCPREFDLYLEAFGLISDFLDH-----QAVIKFFFELETHF 155
Db 257 AEEIATWKEVYTTLAGLYATHACGEHLEAFALLERFSGYREDNIPOLEDVSRFLKRTGF 316
QY 156 SYYPVSGFVAPHQVLSLLQDRVFPTASVMTLDKNFSLTDLIHDLLGHVPLLLHPSFS 215
Db 317 QLRPVAGLLSARDPLASIAFRVFOCTQYIRHASSPMHSPSPDCCHLLGHVPLADRTFA 376
QY 216 EFINNGRLFTKVIKVOALSKKQRTQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYG 275
Db 377 QFSQDIG-----LASLGASDEEIEKLTLS-----WFTVEFLCKONGEVKAYG 420
QY 276 AVLISSPQEGHAFIDNVRLPLELDQIIRLPNTPTPQTLFSIRHF-----DELVELTS 331
Db 421 AGLLSYGGELHLLCEEPERAFDPEAAAVQPYQDYQSYQVSVSEFSFSDAKDLRSYAS 480
QY 332 KLE 334
Db 481 RIQ 483

RESULT 14
TY3H_DROME

TY3H_SCHMA STANDARD; PRT; 508 AA.
ID P18459;
AC 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH)
DE (PROTEIN PALE)
GN TH OR PLE.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90166583; PubMed=2483109;
RA McKameyer W.S.; Quinn W.G.;
RA "Isolation and characterization of the gene for Drosophila tyrosine
hydroxylase";
RN Neuron 2:1167-1175(1989).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14395; AAA62877.1; -
CC EMBL; X76209; CAA53802.1; -
CC PIR; JN0010; JN0010.
CC HSSP; P04177; ITOH.
CC FlyBase; FBgn0005626; ple.
CC InterPro; IPR001273; -
CC Pfam; PF00351; bioterin.H; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis.
FT METAL 338 338 IRON (BY SIMILARITY).
FT METAL 343 343 IRON (BY SIMILARITY).
FT METAL 383 383 IRON (BY SIMILARITY).
SQ SEQUENCE 508 AA; 57861 MW; 2D194E0E8495E66F CRC64;

Query Match 11.6%; Score 219; DB 1; Length 508;
Best Local Similarity 28.6%; Pred. No. 4.3e-10;
Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;

QY 128 DYLEAFGLSD---FLDH-----QAVIKFELETHFSYYPVSGFVAPHOYLSLLQDRYFP 179
DB 258 EYRAAFQRLQEQIEFVETRLPOLQMSDFLRKNTGSLRPAAGLLTARDLASFARIFQ 317
QY 180 IASVMTLDKDNFSLTDLHDLGHVPWLLHPSFSEFFINMGRLFTKVIKVOALPSKK 239
DB 318 SQYVYRVHNSPYHTPEPDSITHELGHMPLADPSPAQFSQETG-----LASLGASD 368
QY 240 QRIOTLQSNLIAVRCFWFVSGLEIENHEGRKAYGAVLISSPQELGHAFID-----291
DB 369 EIEKLSL-----VYWFVVEGLCKHQIKAYGAGLGLSSYGELLHAISDKCEHRAFE 421
QY 292 --NVRLPLE-----LDQIRLPNTSTPQETLFSIRHFD--ELVELT 330
DB 422 PASTAVQYQDQEQYQIYYVAESFEDAKOKFRWSTKSRPEVRFN-PHTERVEVLDVS 480

QY 331 SKLEWMLDQGLLESIPLYN 349
DB 481 DKLETLVHOMNTTEILHLTN 499

RESULT 15
TY3H_SCHMA STANDARD; PRT; 465 AA.
AC O17446;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).
GN TH.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Strigeidae; Schistosomatidae; Schistosomatidae;
OC Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PUERTO RICAN;
RA MEDLINE=98421787; PubMed=97511167;
RA Hamdan F.F.; Ribeiro P.;
RA "Cloning and characterization of a novel form of tyrosine hydroxylase
from the human parasite, Schistosoma mansoni";
RL J. Neurochem. 71:1369-1380(1998).
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + DIHYDROPTERIDINE + O(2) =
CC 3,4-DIHYDROXY-L-PHENYLALANINE + TIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF030336; AAC62256.1; -
CC HSSP; P04177; ITOH.
CC InterPro; IPR001273; -
CC Pfam; PF00351; bioterin.H; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis.
FT METAL 294 294 IRON (BY SIMILARITY).
FT METAL 299 299 IRON (BY SIMILARITY).
FT METAL 339 339 IRON (BY SIMILARITY).
SQ SEQUENCE 465 AA; 54081 MW; 0DE47B4A474665CB CRC64;

Query Match 11.5%; Score 216.5; DB 1; Length 465;
Best Local Similarity 24.9%; Pred. No. 6.1e-10;
Matches 59; Conservative 49; Mismatches 88; Indels 41; Gaps 5;

QY 128 DYLEAFGLSDFLDH-----QAVIKFELETHFSYYPVSGFVAPHOYLSLLQDRYFP 179
DB 214 EYIDGFOLEKYCNVNSSESTIPQLTICEFMHRTSGFIRPVAGLVSPKDFLASLAFRVFQ 273
QY 180 IASVMTLDKDNFSLTDLHDLGHVPWLLHPSFSEFFINMGRLFTKVIKVOALPSKK 239
DB 274 CTQVIRHHSRPMHTPEPDCITHELIGHMPLVNRQFADPSQELG-----LASLGASEET 327
QY 240 QRIOTLQSNLIAVRCFWFVSGLEIENHEGRKAYGAVLISSPQELGHAFIDNVRLPLE 299
DB 328 TRLSLT-----YWFVVEGLCENGETRALGAGIMSSYGELENAP-----367
QY 300 LDQIRLPNTSTPQETLFSIRHFDVELTSLKLEWM-----LDQGLLESIPLYN 349

Mon Jul 2 08:46:44 2001

us-09-438-185-1047.rsp

Page 11

Db 368 SLSVKEFFNINDAAVOYDDVGQKIYFVTESIESMKRELNYINTSGKSTPIYD 424

Search completed: June 29, 2001, 13:56:09
Job time: 258 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:52:51 ; Search time 13.54 Seconds
(without alignments)
538.579 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHYCERTLPKYLKALKL.....ESIPLYNQEKYLSGFVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/ECTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	223	11.8	498	1	US-08-009-075-5
2	220	11.6	497	1	US-08-009-075-4
3	98.5	5.2	1024	2	US-09-091-117-5
4	95	5.0	1226	2	US-08-540-804-12
5	95	5.0	1226	2	US-08-218-265-12
6	95	5.0	1226	3	US-08-321-872-12
7	95	5.0	1226	3	US-08-590-399-12
8	91.5	4.8	872	1	US-08-766-014-2
9	91	4.8	535	3	US-08-369-822C-24
10	91	4.8	535	3	US-08-582-776C-39
11	91	4.8	535	4	US-08-434-831B-36
12	91	4.8	2183	1	US-08-348-891A-7
13	91	4.8	2183	2	US-08-905-817-7
14	90.5	4.8	615	1	US-08-484-105-6
15	90.5	4.8	615	1	US-08-484-106-6
16	89	4.7	392	3	US-08-911-853-35
17	89	4.7	392	4	US-09-479-409-35
18	87.5	4.6	309	3	US-08-996-338-22
19	87.5	4.6	521	3	US-08-996-338-20
20	87.5	4.6	541	1	US-08-604-333-2
21	87.5	4.6	541	3	US-09-110-618-2
22	86.5	4.6	990	2	US-08-392-625-20
23	86.5	4.6	990	2	US-08-466-963A-20
24	86.5	4.6	990	2	US-08-445-193B-15
25	83.5	4.4	1365	6	5194600-4
26	83	4.4	491	2	US-08-942-819-2
27	82	4.3	861	3	US-09-022-875-2

28	81.5	4.3	608	1	US-08-766-014-3	Sequence 3, Appli
29	81.5	4.3	1708	1	US-08-493-092-2	Sequence 2, Appli
30	81.5	4.3	1708	1	US-08-508-836A-2	Sequence 2, Appli
31	81.5	4.3	3056	2	US-08-508-836A-8	Sequence 8, Appli
32	81.5	4.3	3056	2	US-08-629-001A-3	Sequence 3, Appli
33	81.5	4.3	3056	4	US-08-874-266-2	Sequence 2, Appli
34	81.5	4.3	3056	4	US-08-642-274D-3	Sequence 3, Appli
35	81.5	4.3	3056	4	US-08-952-127-3	Sequence 3, Appli
36	81	4.3	323	2	US-08-540-804-18	Sequence 18, Appli
37	81	4.3	323	2	US-08-218-265-18	Sequence 18, Appli
38	81	4.3	323	3	US-08-521-872-18	Sequence 18, Appli
39	81	4.3	323	4	US-08-590-399-18	Sequence 18, Appli
40	81	4.3	474	4	US-08-961-083-60	Sequence 60, Appli
41	81	4.3	1118	2	US-08-724-354D-2	Sequence 2, Appli
42	81	4.3	1118	3	US-09-270-984A-2	Sequence 2, Appli
43	80.5	4.3	377	1	US-07-863-169A-1	Sequence 1, Appli
44	80.5	4.3	377	2	US-08-429-964-1	Sequence 1, Appli
45	80.5	4.3	377	3	US-07-935-087-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-009-075-5
; Sequence 5, Application US/08009075
; Patent No. 5300436
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Menek
; APPLICANT: WU, Jing
; APPLICANT: FILER, David
; APPLICANT: FRIEDHOFF, Arnold J.
; TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE
; HYDROXYLASE AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,075
; FILING DATE: 19930126
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY K.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: GOLDSTEIN-1A
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 498 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-009-075-5

Query Match 11.8%; Score 223; DB 1; Length 498;
Best Local Similarity 26.78%; Pred. No. 9.9e-16;
Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

Qy 112 SSRFLWKS-----YCPRFLLDYLEAFGLSLDFLDH-----QAVIKFFLETHF 155

Db 227 ABEIATREVTTLKGLYATHACREHLEGQLLERYCGYREDSPQLEDVSRFLKERTGF 286
QY 156 SYYPVSGVAPHQVLSLLQDRYFPPIASVMRTLDKNFSLTPDLIHDLLGHVPLLLHPSFS 215
Db 287 QLRPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMPLADRTFA 346
QY 216 EFFINMGRLEFTKVIKQALPSKKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYG 275
Db 347 QFSODIG-----LASIGASDEIEKLSLTS-----VFVTFEGLCKQNGELKAYG 390
QY 276 AVLISSPOELGHAFIDNVRLPLEDOIIRLPNTSTPOETLFSIRHF-----DELVELTS 331
Db 391 AGLSSYGELLHSLSEPEVRAFDPTAAVQPDQOTYQPVYFVSESFNDKDLRNYAS 450
QY 332 KLE 334
451 RIQ 453

RESULT 2
US-08-009-075-4
; Sequence 4, Application US/08009075
; Patent No. 5300436
; GENERAL INFORMATION:
; APPLICANT: GOLDSTEIN, Menek
; APPLICANT: WU, Jing David
; APPLICANT: FRIEDHOFF, Arnold J.
; TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE
; TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,075
; FILING DATE: 19930126
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY K.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: GOLDSTEIN-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5137
; TELEFAX: 202-737-3328
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-009-075-4

Query Match 11.6%; Score 220; DB 1; Length 497;
Best Local Similarity 26.7%; Pred. No. 2.1e-15;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;
QY 112 SRFSFLWKS-----YCPRFELDYLEAFGLLSDFLDH-----QAVIKFFLELTHF 155
Db 226 ABEIATREVTTLKGLYATHACREHLEGQLLERYCGYREDSPQLEDVSRFLKERTGF 285
QY 156 SYYPVSGVAPHQVLSLLQDRYFPPIASVMRTLDKNFSLTPDLIHDLLGHVPLLLHPSFS 215

Db 286 QLRPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMPLADRTFA 345
QY 216 EFFINMGRLEFTKVIKQALPSKKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYG 275
Db 346 QFSODIG-----LASIGASDEIEKLSLTS-----VFVTFEGLCKQNGEVKAYG 389
QY 276 AVLISSPOELGHAFIDNVRLPLEDOIIRLPNTSTPOETLFSIRHF-----DELVELTS 331
Db 391 AGLSSYGELLHSLSEPEVRAFDPTAAVQPDQOTYQPVYFVSESFNDKDLRNYAS 449
QY 332 KLE 334
451 RIQ 452

RESULT 3
US-09-091-117-5
; Sequence 5, Application US/09091117
; Patent No. 6171589
; GENERAL INFORMATION:
; APPLICANT: The University of Melbourne
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,117
; FILING DATE: 12 JUNE 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU96/00803
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN7127
; FILING DATE: 13-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WINNER, Ellen P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: +1 303 499 8080
; TELEFAX: +1 303 499 8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma genitalium
US-09-091-117-5

Query Match 5.2%; Score 98.5; DB 4; Length 1024;
Best Local Similarity 19.9%; Pred. No. 0.091;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;
QY 29 QNSQSLQRAYSTPYSYRIILOKENKEK---QALARKHCISILEFFKNLLFVHLLSLKN 85
Db 564 QOTDSLKNLFSV---IGDILSETNVNKTTHAVKNNELLSLVETASTLKIKHL----- 613
QY 86 QRECSCTDMAVVSTPFPFNRLNWLRLSSRFLSWKSYCPRFFLDYLEAFGLLSDFLDHQA 145
Db 614 -----NVQYKVLVDKFKELKNSFIK-----ELNFFPDTKDITPT 647

QY 146 IK--PFELETH-----FSYVPVSGFVAPHOYLSLLQDRYFFPIASVMTLDKDNF 192
Db 648 IKVLESENYTLKKYENEGFGYHWAFIVPGTFNSAENTFSAL-----DKT 698
QY 193 SLTPDLIHLHGHVPLWLPHPSEFFINMGRLET-----KVIEKVOALPSKKQRIQT 244
Db 699 KSTRDLFADML--FGKLSVSDSDSFIKINGSETLKYHGDNLNLLPNYHSLITKNVGYOI 757
QY 245 LQSNLIAVRCFWFFVESGLIENHEGRKAYGAVLSSPOELG-----286
Db 758 VVNFHIDARLLTABLQNTVSNPK-----PVIKSPVELSKSLFEVWKTFIENSVMQI 810
QY 287 ---HAFIDNVRVPLEDQIIRLPFNFTSTPQETLFSIRHFD 324
Db 811 LKKEYTFKDNLFKFFPKAGDSRSLFELDKPDQVIPPFAVD 852
RESULT 4
US-08-540-804-12
; Sequence 12, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/540,804
; APPLICATION NUMBER: US/08/540,804
; FILING DATE: 11-OCT-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/521,872
; FILING DATE: 21-AUG-1995
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-NAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03A2
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 1226 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-540-804-12
Query Match 5.08; Score 95; DB 2; Length 1226;
Best Local Similarity 18.8%; Pred. No. 0.29;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;
QY 12 YILKIALKRLQSLSLFFQNSQRAYSTPYSYRII-----LOKENKEQALARKHCI 65

Db 133 YILE---KLIFDMTHYNDSQL--RTWKROIISYFLKLLGNCYSRLINKE---IPHHWLV 65
QY 66 SILEFPKNLF---VHLLSLSKNOREGCGCTMAVST-----PPF-----102
Db 66 EFINKMENEFPLSLHILMIFWMDICOIDTINAPVAATITTSQKEPFFLVTKITDMLHK 125
QY 103 -----NRLNWLRLSSRFSL-----WKS 120
Db 126 YIYSSKSMINDENYIINDIKNNKIKLNLKILSLILKIFQESLEVFIFPTSNWEI 185
QY 121 YCPRFF-----LDYLEAFGLSDFDLHQAVIKFFTELETHFSYYPVSGFVAP 166
Db 186 YKPLLEFIVSNADTNNQSDMKKLELISYRNESLKNNSIR-----NVIMSASNA 236
QY 167 HQVLSLQDRYFPIASVM-----RTLDKDN-----FSLTPDLIHLHGHVP 207
Db 237 DFQLTIVTCKQPKLSCIQLNCIDTQFTKLLD-DNPTFDWPTYVDQNLTMHKIQLIL 295
QY 208 WLLHPS--PSEFFINMGRFLTUVIEKVOALPSKKORIOTLOSNI--IAIVRCFWFTVSG 263
Db 296 WSHPSRQFDHYESN-----QLVAKLLLL-----RINSTDEDLHEFQIEDAINSLVFO- 343
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFID-- 291
Db 344 LAKNSAQRRVSYVMPSLYRLNLILITYGIKVPYIRKLLISSGLLYLQDSNDRKFVHVQ 403
QY 292 ---NVRVPLEDQIIRLPFNFTSTPQETLFSIRHDELVELTSL-----333
Db 404 LLINKISPLMKSOYNMVLNRVMEYDKVEIFNFDQVEITEQIKRILSNIDITNLQLS 463
QY 334 -----EWMLD---OGLLESIPLYNQEKYLSGFVLC 361
Db 464 KTPLSIKIMVAEWYLSHLCSGILSV---NRTVLLKIFKIFC 502
RESULT 5
US-08-218-265-12
; Sequence 12, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/218,265
; FILING DATE: 25-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1226 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-218-265-12

Query Match 5.0%; Score 95; DB 2; Length 1226;
Best Local Similarity 18.8%; Pred. No. 0.29;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKRLQSLFQNSQSLQRAYSTPYSYRII-----LOKENKEQALARKKI 65
DB 13 YILE---KLIFDMNHYNDSQOL-RTWKROIYSYFLKLGNCYSRLINKE---IFHHLV 65
QY 66 SILEFFKNLLF---VHLLSLSKNOREGSDMAVST-----PFF----- 102
DB 66 EFINKMENFEPLSLHILMIFWNDICQIDTNPVAAITSSQKEPFLVTKITDMLLHK 125
QY 103 -----NRNLWRLSSRFSL-----WKS 120
DB 126 YIVSSSKSMINDENYIINDIKNNKIKILNLIKLSLILKIFQOSLEVFIFPTSWEI 185
QY 121 YCPREF-----LDYLEAFGLSDPLDHOAVIKFEFELETHFSYYPVSGFVAP 166
DB 186 YKPLLEIVSNADTNQNSDKMKKLELISYRNESLKNSSIR-----NVIMSASNAN 236
QY 167 HOYLSLLODRFFPIASVM-----RTLDKN-----FSLTPDLIHLDLGHVP 207
DB 237 DFQLTIVTCKQFPKLSQIQLNCIDTOFTKLLD-DNPTFDMPTVYDQNPMTMKIIL 295
QY 208 WLLHPS--FSEFFINMGRLETKVIEKVALPSKKQRIOTLOSNI--IAIVRCFWFTVESG 263
DB 296 WSIHPSRQFDHYESN-----QLVAKLLL-----RINSTDEDLHEFDIAWSLVFQ- 343
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFID-- 291
DB 344 LAKNFSACKRVVSYMPSLYRLNLITLITYGIKVPYIRKLSSGLLYLQDSNDKFEVHQ 403
QY 292 ---NVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVELTSKL----- 333
DB 404 LLINKISPLMKSOYNKVLNRYMVEYDKVFEYFNFDQVLEITEQIKMRILSDITNLQLS 463
QY 334 -----EWMLD---OGLESPLYNQEKYLSGFEVLC 361
DB 464 KTPLSIKIMVAEWYLSHLCSGILSSV---NRTVLLKIFKIFC 502

RESULT 6

US-08-521-872-12
; Sequence 12, Application US/08521872
; Patent No. 6015682
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Willita Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/521,872

; FILING DATE: 31-AUG-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-521-872-12

Query Match 5.0%; Score 95; DB 3; Length 1226;

Best Local Similarity 18.8%; Pred. No. 0.29;

Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKRLQSLFQNSQSLQRAYSTPYSYRII-----LOKENKEQALARKKI 65
DB 13 YILE---KLIFDMNHYNDSQOL-RTWKROIYSYFLKLGNCYSRLINKE---IFHHLV 65
QY 66 SILEFFKNLLF---VHLLSLSKNOREGSDMAVST-----PFF----- 102
DB 66 EFINKMENFEPLSLHILMIFWNDICQIDTNPVAAITSSQKEPFLVTKITDMLLHK 125
QY 103 -----NRNLWRLSSRFSL-----WKS 120
DB 126 YIVSSSKSMINDENYIINDIKNNKIKILNLIKLSLILKIFQOSLEVFIFPTSWEI 185
QY 121 YCPREF-----LDYLEAFGLSDPLDHOAVIKFEFELETHFSYYPVSGFVAP 166
DB 186 YKPLLEIVSNADTNQNSDKMKKLELISYRNESLKNSSIR-----NVIMSASNAN 236
QY 167 HOYLSLLODRFFPIASVM-----RTLDKN-----FSLTPDLIHLDLGHVP 207
DB 237 DFQLTIVTCKQFPKLSQIQLNCIDTOFTKLLD-DNPTFDMPTVYDQNPMTMKIIL 295
QY 208 WLLHPS--FSEFFINMGRLETKVIEKVALPSKKQRIOTLOSNI--IAIVRCFWFTVESG 263
DB 296 WSIHPSRQFDHYESN-----QLVAKLLL-----RINSTDEDLHEFDIAWSLVFQ- 343
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFID-- 291
DB 344 LAKNFSACKRVVSYMPSLYRLNLITLITYGIKVPYIRKLSSGLLYLQDSNDKFEVHQ 403
QY 292 ---NVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVELTSKL----- 333
DB 404 LLINKISPLMKSOYNKVLNRYMVEYDKVFEYFNFDQVLEITEQIKMRILSDITNLQLS 463
QY 334 -----EWMLD---OGLESPLYNQEKYLSGFEVLC 361
DB 464 KTPLSIKIMVAEWYLSHLCSGILSSV---NRTVLLKIFKIFC 502

RESULT 7

US-08-590-399-12
; Sequence 12, Application US/08590399
; Patent No. 6214588
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor

Db 310 TLQD-----NNIAFYPLIVEGEIKKIAFGKDTVYIDVFQTE-----DL 351
QY 184 MRLDKDNFSLTPDLIDL-----GHVPWLLHPSEFFINWGLFT 226
Db 352 KEIFKEDEFETHEIKDFLVRISYKIGIECKSKYIDTAVMAYLLMPSESNT--DLDRVLK 409
QY 227 KVIKQALPS-----KKORIOTLOSNIU-----IVRCFW----- 257
Db 410 KVL-KVD-VPSYEGIFGKGKDKKKEIDENILADYICRVCYILFDLKEKLMNFIERMDM 467
QY 258 ----FTVESGLIENHEGRKAYGAVLISSP-QELGHAFIDNVRVLPLELDQIIR--LPFN 309
Db 468 KLLLEIEMPLEVEVLKMEYSVGTLDKEVLKLSQKIDDRIGEI--LDKIYKEAGYQFN 524
QY 310 TSTPOE-TLFSIRHFDLVELTSLKLEWMLDQGLLESIPLYN-----OKEY 353
Db 525 VNSPKQLSEFLKLNLPVKKTKGTGYSTDSEVLQVLPYNDIVSDIIEYRQLTKLKSTY 584
QY 354 LSGF 357
Db 585 IDGF 588

RESULT 9
US-08-369-822C-24
; Sequence 24, Application US/08369822C
; Patent No. 6015860
; GENERAL INFORMATION:
; APPLICANT: Lipkin, W. I.
; APPLICANT: Briese, Thomas
; APPLICANT: Kliche, Stefanie
; APPLICANT: Schneider, Patrick A.
; APPLICANT: Stitz, Lothar
; APPLICANT: Schneemann, Anette
; TITLE OF INVENTION: Borna Disease Viral Sequences,
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
; TITLE OF INVENTION: System Diseases
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 865 South Figueroa Street, 29th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017-2571
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
; SOFTWARE: ASCII (DOS) TEXT
; CURRENT APPLICATION DATA:
; FILING DATE: 06-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Churchill, Margaret A. (Ph.D.)
; REGISTRATION NUMBER: 39,944
; REFERENCE/DOCKET NUMBER: 1279-194XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213/892-9200
; TELEFAX: 213/680-4518
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-369-822C-24

Query Match 4.8%; Score 91; DB 3; Length 535;
Best Local Similarity 20.6%; Pred No, 0.22; Indels 136; Gaps 18;
Matches 78; Conservative 46; Mismatches 119;
QY 57 QALARKICISILEFFKNLLFVILLSLKNQREGGCTDMVVSTPFFNRNLWYLLSSRFS 116
Db 1 EGTHTOCVDNWKSPAGVAF-----GCFMPLSLDSD--LTMYLKDALAALQR 46
QY 117 LKWSYCPREFLDYLEAGLLSFLDQHAVIKVFELETHFSYYPVSGFVAPHQYLSLLQDR 176
Db 47 EMDSVYKPEFLRYDPKGTGS-----RRLDVFLNDSFDPYDVMYVVSAGY---LHDP 98
QY 177 YFTASVMRTLDKDNFSTPDLIHLHGVFWLLHPSEFFINWGLFTKVIKQALP 236
Db 99 EFNL-----SYSLOEKEIKE-----TGRFAKMTYKRA-- 127
QY 237 SKQRIQTLOSNIU-----IVRCFWFTVESG-----LIENHEGRKAY 274
Db 128 ----COVIAENLISNGIKYKPDNGMAKDEQDLTKALHTLAVSGVPKDKESHG--- 178
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPENTSTPOE----- 315
Db 179 GPVLKYSRSPVHTSTRNVRAAKGFPGVQVIRQDDTDHPENMEAYETVSATITDLAK 238
QY 316 -----TLFSIRHFDLVELTSLKLEW-----LDQGLL-----ESIPLY-- 348
Db 239 YCLNRYETISLFAQR-LNEIYGLPSFFQWLHKLRLTSVLVSDPHCPDLDHAHIELYKV 297
QY 349 -NOE---KY-LSGFEVLQ 362
Db 298 PNDQIFIKYPMGGIEGYCQ 316
RESULT 10
US-08-582-776C-39
; Sequence 39, Application US/08582776C
; Patent No. 6077510
; GENERAL INFORMATION:
; APPLICANT: Lipkin, W. I.
; APPLICANT: Briese, Thomas
; APPLICANT: Kliche, Stefanie
; APPLICANT: Schneider, Patrick A.
; APPLICANT: Stitz, Lothar
; APPLICANT: Schneemann, Anette
; TITLE OF INVENTION: Borna Disease Viral Sequences,
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
; TITLE OF INVENTION: System Diseases
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 865 South Figueroa Street, 29th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017-2576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: WINDOWS NT
; SOFTWARE: ASCII DOS TEXT
; CURRENT APPLICATION DATA:
; FILING DATE: 04-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 06-JAN-1995
; PRIOR APPLICATION NUMBER: US 08/369,822
; FILING DATE: 04-MAY-1995
; PRIOR APPLICATION NUMBER: US 08/434,831
; ATTORNEY/AGENT INFORMATION:
; NAME: Churchill, Margaret A.
; REGISTRATION NUMBER: 39,944

REFERENCE/DOCKET NUMBER: 1279-194C2
TELEPHONE: 213/892-9200
TELEFAX: 213/680-4518
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 535 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-582-776C-39

Query Match 4.8%; Score 91; DB 3; Length 535;
Best Local Similarity 20.6%; Pred. No. 0.22;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;
QY 57 QALARKKICISILEFFKNLLFVHLLSLKNOEGCSTDMVSTPFNNRLWYRLSSRFS 116
D 1 EGLTHQCCVDMNKSFAGVKF-----GCFMPLSLDSD--LTMYLKDKKALAAALOR 46
QY 117 LKMSYCPREFDYLEAFGLSLDLDHQAVIKFLETHFSYYPVSGFVAPHOYLSLQDR 176
D 47 EWDVYVPKREFLRYPDPKGTGS-----RLVDVFLNDSSFPDYDVMYVSGAY---LHDP 98
QY 177 YFPIASVMTLKDKNFSLTPDLIHDLGHVPMLLHPSFEFFINMGRLETKVIEKVOALP 236
D 99 EFNLS-----SYSLQKEIKE-----IVRCFWFTVESG-----LIENHEGRKAY 127
QY 237 SKKORIQTLOSNLIA-----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHARG 178
D 128 -----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHARG 178
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPOE----- 315
D 179 GPKVLTYSRSPVHTSTRNVRRAAKGFIGFPQVIRQDQDTHPENMEAYETVSAFTTDLKK 238
QY 316 -----TLFSIRHDFELVELTSLKLEWM---LDQGLL-----ESIPLY-- 348
D 239 YCLNWRVETISLFAQR-LNEIYGLPSFFQWLHKRLTSLVLYVSDPHCPDLDLDAHIPLYK 297
QY 349 -NOE---KY-LSGFEVLQ 362
D 298 PNDQIFIKYPMGGIEGYCQ 316

RESULT 11
US-08-434-831B-36
Sequence 36, Application US/08434831B
Patent No. 6113905
GENERAL INFORMATION:
APPLICANT: Lipkin, W. I.
APPLICANT: Briese, Thomas
APPLICANT: Kliche, Stefan
APPLICANT: Schneider, Patrick A.
APPLICANT: Stitz, Lothar
APPLICANT: Schneemann, Anette
TITLE OF INVENTION: Borna Disease Viral Sequences
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
TITLE OF INVENTION: System Diseases
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski, L.L.P.
STREET: 865 South Figueroa Street, 29th Floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017-2571
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
SOFTWARE: ASCII (DOS) TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,831B
FILING DATE: 04-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/369,822
FILING DATE: 06-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Churchill Margaret A. (Ph.D.)
REGISTRATION NUMBER: 39,944
REFERENCE/DOCKET NUMBER: 1279-194C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213/892-9200
TELEFAX: 213/680-4518
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 535 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-434-831B-36

Query Match 4.8%; Score 91; DB 4; Length 535;
Best Local Similarity 20.6%; Pred. No. 0.22;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;
QY 57 QALARKKICISILEFFKNLLFVHLLSLKNOEGCSTDMVSTPFNNRLWYRLSSRFS 116
D 1 EGLTHQCCVDMNKSFAGVKF-----GCFMPLSLDSD--LTMYLKDKKALAAALOR 46
QY 117 LKMSYCPREFDYLEAFGLSLDLDHQAVIKFLETHFSYYPVSGFVAPHOYLSLQDR 176
D 47 EWDVYVPKREFLRYPDPKGTGS-----RLVDVFLNDSSFPDYDVMYVSGAY---LHDP 98
QY 177 YFPIASVMTLKDKNFSLTPDLIHDLGHVPMLLHPSFEFFINMGRLETKVIEKVOALP 236
D 99 EFNLS-----SYSLQKEIKE-----IVRCFWFTVESG-----LIENHEGRKAY 127
QY 237 SKKORIQTLOSNLIA-----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHARG 178
D 128 -----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHARG 178
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPOE----- 315
D 179 GPKVLTYSRSPVHTSTRNVRRAAKGFIGFPQVIRQDQDTHPENMEAYETVSAFTTDLKK 238
QY 316 -----TLFSIRHDFELVELTSLKLEWM---LDQGLL-----ESIPLY-- 348
D 239 YCLNWRVETISLFAQR-LNEIYGLPSFFQWLHKRLTSLVLYVSDPHCPDLDLDAHIPLYK 297
QY 349 -NOE---KY-LSGFEVLQ 362
D 298 PNDQIFIKYPMGGIEGYCQ 316
RESULT 12
US-08-348-891A-7
Sequence 7, Application US/08348891A
Patent No. 5654136
GENERAL INFORMATION:
APPLICANT: SASAKI, Keiko
APPLICANT: MORI, Takayuki
APPLICANT: MAKINO, Satoshi
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/348,891A
; FILING DATE: 25-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 10-MAR-1992
; APPLICATION NUMBER: US 07/848,400
; PRIOR APPLICATION DATA:
; FILING DATE: 14-OCT-1991
; APPLICATION NUMBER: JP 3-293625
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-348-891A-7

```

```

Query Match 4.8%; Score 91; DB 1; Length 2183;
Best Local Similarity 20.6%; Pred. No. 1.9;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARKKICISILEFFKNLLFVHLLSLSKNQREGCGTDMVSTPTFFNRLMYRLSSRFS 116
DB 429 EGLTHQOCVDNNKWSFAGVKF-----GCFMPLSLDSD--LTMYLKDKALAAALQR 474
C 117 LKWSYCPREFLDYLEAFGLSLDFLDHQAVIKFFETHFYSYVPSGVFVAPHQYLSLQDR 176
DB 475 EWDVYPKFELRYDPPKGTGS-----RRLVDVFLNDSFDPYDVMYVVSAGY---LHDP 526
DB 177 YFPIASVMRTLDKDNFSLTDPDLIHDLLGHVPWLLHPSFSEFFINMGRLETKVIEKVOALP 236
DB 527 EFNL-----SYSLKEKE-----TGRLEFAKMTYKRA-- 555
QY 237 SKKORIOTLOSNLIA-----IVRCFWFTVESG-----LIENHEGRKAY 274
DB 556 -----CQVIAENLISNGIKYFKDNGMAKDEHDLTKALHTLAVSGVPKDLKESHG-- 606
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315
DB 607 GVLKTSRSRSPVHTSTRNRAAKGFTGFQVIRQODQDHPENMEAYETVSATITDLKK 666
QY 316 -----TLFSIRHFDLVELTSKLEWM--LDQGL-----ESIPLY-- 348
DB 667 YCLNWRVETISLFAQR--LNEIYGLPSFFQWLHKRLTSLVLYSDPHCPDLDHAIPLYKV 725
QY 349 -NQE---KY--LSGFEVLQC 362
DB 726 PNDQIFIKYPMGGIEGYC 744

RESULT 13
US-08-905-817-7

```

```

; Sequence 7, Application US/08905817
; Patent No. 5824777
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE.
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,817
; FILING DATE: 04-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,891
; FILING DATE: 25-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-905-817-7

```

```

Query Match 4.8%; Score 91; DB 2; Length 2183;
Best Local Similarity 20.6%; Pred. No. 1.9;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARKKICISILEFFKNLLFVHLLSLSKNQREGCGTDMVSTPTFFNRLMYRLSSRFS 116
DB 429 EGLTHQOCVDNNKWSFAGVKF-----GCFMPLSLDSD--LTMYLKDKALAAALQR 474
QY 117 LKWSYCPREFLDYLEAFGLSLDFLDHQAVIKFFETHFYSYVPSGVFVAPHQYLSLQDR 176
DB 475 EWDVYPKFELRYDPPKGTGS-----RRLVDVFLNDSFDPYDVMYVVSAGY---LHDP 526
QY 177 YFPIASVMRTLDKDNFSLTDPDLIHDLLGHVPWLLHPSFSEFFINMGRLETKVIEKVOALP 236
DB 527 EFNL-----SYSLKEKE-----TGRLEFAKMTYKRA-- 555
QY 237 SKKORIOTLOSNLIA-----IVRCFWFTVESG-----LIENHEGRKAY 274
DB 556 -----CQVIAENLISNGIKYFKDNGMAKDEHDLTKALHTLAVSGVPKDLKESHG-- 606
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315

```

Db 607 GPKYTSRSPVHTSTRNVRRAAKGFIGPQVIRQDQDTHDPENMEAYETVSFAITTDLKK 666
Qy 316 -----TLFSIRHDELVELTSKLEWM---LDQGLL-----ESIPLY-- 348
Db 667 YCLNWRYETISLFAQR-LNEYIGLPSFPQWLHKKRLTSLVYSDPHCPDLDHAIPLYKV 725
Qy 349 -NOE---KY-LSGFEVLQ 362
Db 726 PNDQIFIKYPMGGIEGYCQ 744

RESULT 14
US-08-484-105-6

; Sequence 6, Application US/08484105
; Patent No. 5589341

; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce

; APPLICANT: BELL, Stephen P

; APPLICANT: KOBAYASHI, Ryuji

; APPLICANT: RINE, Jasper

; APPLICANT: FOSS, Margit

; APPLICANT: McNALLY, Francis J

; APPLICANT: LAURENSEN, Patricia

; APPLICANT: HERSKOWITZ, Ira

; APPLICANT: LI, Joachim J

; APPLICANT: GAVIN, Kimberly

; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,105

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard Aron

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO

; TELEPHONE: (415) 494-8700

; TELEFAX: (415) 494-8771

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 615 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-484-105-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;
Best Local Similarity 20.7%; Pred. No. 0.31;
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

Qy 24 LSLFTQNSQSLQRAYSTPYIS-----YVRILQKNEKQALARKHCISILEFFKNLLF 76

Db 326 MSYFFQNAFVS---FIDPVNVDFLNDYLKIL-----SRCPTFMFEVGLIK 369

Qy 77 VH-----LILSLKNOREGCS---TDMNVVSTP-----FFNENLWYLLSSRFLSKSYC 122

Db 370 QHAPADEILSLTNKRNKGLIEFFVEFLVRENPINGHAKFVARELEELNITNFNLIELY- 428
Qy 123 PREFDLYLEAFGLLDLFDHOAVIKFFFELETHFSYYPV-----SGFVAPH-- 167
Db 429 -----HNLICKLSDYLDRWSACKYKDLRHF--PIDTIFQELFTLDRSGLLTQSIF 480
Qy 168 -QYLSLLQDRYFPPIASVMTLQKNF-SLTPDL---IHDLLGHVPWLLHP-----SFSEF 217
Db 481 PSYKSNIEDNLLSWEQVLPDLKENYDTLSGDLDKINAPVLGOLFPLYREANWTINIDF 540
Qy 218 FINMGRLFTK--VIEKVQALPSKKQRTQTLQS-----NLIAIVRCFWFTVESGLIENH 268
Db 541 YIAFRETLPKEILNFIKRDYSNTKLELAETPDADFVKVALILFMOAIFAFENMGLIK-F 599
Qy 269 EGRKAYGAV 277
Db 600 QSTKSYDLV 608

RESULT 15

US-08-484-106-6

; Sequence 6, Application US/08484106

; Patent No. 5614618

; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce

; APPLICANT: BELL, Stephen P

; APPLICANT: KOBAYASHI, Ryuji

; APPLICANT: RINE, Jasper

; APPLICANT: FOSS, Margit

; APPLICANT: McNALLY, Francis J

; APPLICANT: LAURENSEN, Patricia

; APPLICANT: HERSKOWITZ, Ira

; APPLICANT: LI, Joachim J

; APPLICANT: GAVIN, Kimberly

; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,106

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard Aron

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO

; TELEPHONE: (415) 494-8700

; TELEFAX: (415) 494-8771

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 615 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-484-106-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;
Best Local Similarity 20.7%; Pred. No. 0.31;
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 521
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-996-338-20

Query Match 4.6%; Score 87.5; DB 3; Length 521;
Best Local Similarity 21.3%; Pred. No. 0.5;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;
QY 9 DPXYI-----LKIALKRLQSLFLFONQSLOLAYSTPYSYVRIILQENKEKQALAHKC 64
Db 15 EPPYLKHCSCSLAHEITTTKSWKSSGQEHVELNPRSSRIAL-----HDC 62
QY 65 ISLEFF-----KNLLFVHLLSLKNORECCSTDMVVS-----TPFFN-- 103
Db 63 --VLEFWPELVNDTGSFFOMKNYTKQKLNVRNKHSCFTEQVTSKIVEVKKFFOIT 120
QY 104 -RNLWYRLSSRFLSKSYCPREFLDYLEAFGLLSDFLDHQAVIKFPELETHFSYYPVSG 162
Db 121 CENSYIOTLVNSTSLYKN-CKLLLENKK-----NPTIKKNAEFEDQ-GYSCVH 168
QY 163 FVAPHQYLSLQDRYFFIASVMRTLDKDNFSLTPDLIHDLLGHVPLWHPSPSEFFINMG 222
Db 169 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVPVLGPKLNHVAVELG 211
QY 223 RLFTKVIKQVQALPSKKQRIQTQLSNLIAIVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280
Db 212 K---NVLNCSALLNEEDVI-----YWMFGEENGSDPNIEEKE-----MRIM 251
QY 281 SPOELGHAFIDNVRVLPLE 299
Db 252 TPEGKWAH-----SKVLRIE 266

RESULT 20
US-08-604-333-2
Sequence 2, Application US/08604333
Patent No. 5776731
GENERAL INFORMATION:
APPLICANT: Parnet, Patricia et al.
TITLE OF INVENTION: Receptor Designated 2F1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathryn A. Anderson, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
FILING DATE: 21-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Anderson, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2619
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 541 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-604-333-2

Query Match 4.6%; Score 87.5; DB 1; Length 541;
Best Local Similarity 21.3%; Pred. No. 0.53;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;
QY 9 DPXYI-----LKIALKRLQSLFLFONQSLOLAYSTPYSYVRIILQENKEKQALAHKC 64
Db 34 EPPYLKHCSCSLAHEITTTKSWKSSGQEHVELNPRSSRIAL-----HDC 81
QY 65 ISLEFF-----KNLLFVHLLSLKNORECCSTDMVVS-----TPFFN-- 103
Db 82 --VLEFWPELVNDTGSFFOMKNYTKQKLNVRNKHSCFTEQVTSKIVEVKKFFOIT 139
QY 104 -RNLWYRLSSRFLSKSYCPREFLDYLEAFGLLSDFLDHQAVIKFPELETHFSYYPVSG 162
Db 140 CENSYIOTLVNSTSLYKN-CKLLLENKK-----NPTIKKNAEFEDQ-GYSCVH 187
QY 163 FVAPHQYLSLQDRYFFIASVMRTLDKDNFSLTPDLIHDLLGHVPLWHPSPSEFFINMG 222
Db 188 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVPVLGPKLNHVAVELG 230
QY 223 RLFTKVIKQVQALPSKKQRIQTQLSNLIAIVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280
Db 231 K---NVLNCSALLNEEDVI-----YWMFGEENGSDPNIEEKE-----MRIM 270
QY 281 SPOELGHAFIDNVRVLPLE 299
Db 271 TPEGKWAH-----SKVLRIE 285

RESULT 21
US-09-110-618-2
Sequence 2, Application US/09110618
Patent No. 6050918
GENERAL INFORMATION:
APPLICANT: Parnet, Patricia et al.
TITLE OF INVENTION: Receptor Designated 2F1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathryn A. Anderson, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
FILING DATE: 09/09/110,618
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/604,333
FILING DATE: 21-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Anderson, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2619
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 541 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-110-618-2

Query Match 4.6%; Score 87.5; DB 3; Length 541;
Best Local Similarity 21.3%; Pred. No. 0.53;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

QY 9 DPXYI---LKALKRQSLSLFQNSQSLRAYSTPYSYRILQRENKQALARKHC 64
DB 34 EPYFKHCSCSLAHEIETTKSWKSSGQEHVELNPRSSRIAL-----HDC 81
QY 65 ISLLEFF-----KNLLFVHLLSLSKNORCGCTDMAVVS-----TPCFN-- 103
DB 82 --VLEFPVELNDGTSYFFQMKNTQKWLNVIRRNKHSCTERQVTSKIVEVKKFOIT 139
QY 104 -RMLWTRLSSRSLWSYCPREFDYLEAFGLSDFLDQAVIKFLETHSYSPVSG 162
DB 140 CENSYOTLVNSTSLYKN-CKKLLNNK-----NPTIKKNAEFDO-GYISCVH 187
163 FVAPHQYLSLQDRYFPFIASVMRTLDKNFSLTPDLIHLGHVPLHPSFSEFFINMG 222
DB 188 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVPLGPKLNHVAVELG 230
QY 223 RLFTKVIKQVQALPSKQRIOTLOSNIATVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280
DB 231 K---NVLNCSALLNEEDVI-----YWMFGENGSDPNIHEEKE---MRIM 270
QY 281 SPOELGHAFIDNVRLPLE 299
DB 271 TPBGKWAH---SKVLRIE 285

RESULT 22

US-08-392-625-20
Sequence 20, Application US/08392625

Patent No. 5837485

GENERAL INFORMATION:

APPLICANT: Entian, Karl-Dieter
APPLICANT: G tz, Friedrich
APPLICANT: Schnell, No. 5837485bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Gerhard
APPLICANT: Rosenstein, Ralf
APPLICANT: Kaletta, Cortina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
APPLICANT: Kellner, Roland

TITLE OF INVENTION: Biosynthetic Process For The Preparation

OF CHEMICAL COMPOUNDS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/392,625

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/876,791

FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.0980002
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-392-625-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;

Best Local Similarity 23.1%; Pred. No. 1.7;

Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;

QY 3 YCERLDPKYILKIAL-KLRQSLSLFQNSQSLRAYSTPYSY-----RIIL 49
DB 601 FCPRIYKNILKATWKINSEM---FSETENWLNRFATIRKWHIPKDVIIAFGDNELL 657
QY 50 OKENKE-----KQALARKHCISILEFFKNLLFVHLLSLSKNORCGCTDMAVSTPFF-- 102
DB 658 NLNDKHLIILKELKHGRIRILESF-----INESNNE-----MLEIYTPLYKK 703
QY 103 -----NRNLWYRLSSRFSLW-----KSYCPREFDYLEAFGL-----LSDFLDHQ 143
DB 704 TSLKEQSFITPKNRKNHFNKDFWSIHLSIPKTYODNFIQDYLPLFTTELKVNFINKF 763
QY 144 AVIKFELETHSYSPVSGFVAPHQYLSLQ--DRYFPFIASVMRTLDKNFSLTPDLI-H 200
DB 764 FYIKKEDEDFIK-----LRLREDEDSQIYSFIKNW-KDYCLLNSELYDY 809
QY 201 DLIGHVP 207
DB 810 SIVDYVP 816

RESULT 23

US-08-466-961A-20

Sequence 20, Application US/08466961A

Patent No. 5843709

GENERAL INFORMATION:

APPLICANT: Entian, Karl-Dieter
APPLICANT: G tz, Friedrich
APPLICANT: Schnell, No. 5843709bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Gerhard
APPLICANT: Rosenstein, Ralf
APPLICANT: Kaletta, Cortina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
APPLICANT: Kellner, Roland

TITLE OF INVENTION: Biosynthetic Process for the Preparation of

CHEMICAL COMPOUNDS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, NW

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

us-09-438-185-1047.rai

Mon Jul 2 08:46:40 2001

QY 123 -----PREFLD-----YLEAFGLLSDFLDH--- 142
Db 228 NGRMSSIPOLKKILYTPKREILVGADNKKQLHDLPEPELRELKLRVTSLISEFYQYKKD 287
QY 143 -----QAVIKFF-----ELETHFSYYPVSGF----- 163
Db 288 ITATLNFTRKSIVNPFLLSKOLIKVSSVKNKIITSNEELNSKGFYNNMLGLYINGQNWKI 347
QY 164 --VAPHOYLSLQDRYFFPIASVMRTDK-----DNFSLTPDLIHDLLGHVPWL--- 209
Db 348 TSLTPYNLLTALKTEYQSLLKITNLQLELEPSKCILDSKFLNKFQFSLGKLNLOPIK 407
QY 210 --LH--PSSEFFINMGR-----FTKVIKVOAL--PSKKORIOTLOSNIATVRCFWF 258
Db 408 MDLHTIPGSESVIYFENDIESDPQYDELVNSVQAFDFKSKFGELPEIKONWSEII---F 463
QY 259 TVESGLIENHEGRKAYG-----AVLISS---POELGHAFIDNVRVLPLELDQIIRLPENTS 311
Db 464 VIDFARLEDESEVKEALGGLVRANVVSOGYTPQVRG-----LLPFSSD 505
312 TPQETLFSIRHF-----DELVELTSKLEWML 337
506 SDKSVNKKIYELKNSTDNLTELKSFLETML 535

Search completed: June 29, 2001, 13:56:34
Job time: 223 sec

THIS PAGE BLANK (USPTO)

Sequence	Strd Orig	ZScore	escore Len	Documentation
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-009-075-3 +		220.00	441.78	5.3e-17
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-08-843-417-9 -		101.50	167.30	0.1032
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-540-804-11 +		95.00	157.84	0.3475
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-218-825-11 +		95.00	157.84	0.3475
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-521-872-11 +		95.00	157.84	0.3475
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-08-509-399-11 +		95.00	157.84	0.3475
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-559-260-1 -		94.00	161.41	0.2198
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-766-014-23 +		91.50	155.44	0.4726
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-766-014-1 -		91.50	146.65	1.46
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-348-891A-1 +		91.00	131.84	9.75
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-905-817-1 +		91.00	131.84	9.75
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-484-105-5 +		90.50	152.88	0.6562
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-484-106-5 +		90.50	152.88	0.6562
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-09-356-952-12 +		90.50	118.12	56.65
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-09-434-774-9 +		89.50	160.78	0.2382
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-596-111B-1 +		89.50	158.69	0.3117
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-911-853-34 +		89.00	159.97	0.2643
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-09-479-409-34 +		89.00	159.97	0.2643
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-911-853-29 +		89.00	136.17	20.17
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-09-479-409-29 +		89.00	136.17	20.17
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-906-338-6 +		87.50	159.65	0.2754
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-996-338-10 +		87.50	158.91	0.3030
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-996-338-1 +		87.50	153.13	0.6358
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-996-338-7 +		87.50	152.68	0.6733
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-604-333-1 +		87.50	152.63	0.6773
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-09-110-618-1 +		87.50	152.63	0.6773
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-392-625-16 +		86.50	129.50	13.17
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-466-961A-16 +		86.50	129.50	13.17
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-645-193B-18 +		86.50	129.50	13.17
/cgn2_6.ptodata/2/ina/5A.COMB.seq:US-08-599-252-90 -		84.50	139.46	3.67
/cgn2_6.ptodata/2/ina/PCTUS.COMB.seq:PCT-US96-06352-90 -		-	84.50	139.46
/cgn2_6.ptodata/2/ina/PCTUS.COMB.seq:PCT-US96-06583-90 -		-	84.50	139.46
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-09-318-448-12 +		84.50	137.25	4.87
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-837-593-1 +		84.00	145.60	1.67
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-540-804-17 +		83.50	143.98	2.05
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-218-265-17 +		83.50	143.98	2.05
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-08-521-872-17 +		83.50	143.98	2.05
/cgn2_6.ptodata/2/ina/6B.COMB.seq:US-08-590-399-17 +		83.50	143.98	2.05
/cgn2_6.ptodata/2/ina/5B.COMB.seq:US-08-942-819-1 +		83.00	136.60	5.30
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-09-022-875-1 -		82.00	134.24	7.70
/cgn2_6.ptodata/2/ina/6A.COMB.seq:US-09-022-875-1 -		82.00	134.24	7.70

```

137 erAspPheLeuAspHis.....GlnAlaVal 145
    :||| :||| :||| :||| :||| :|||
795 AGCGCTTACGGCTTACCGGGAAGACAATATCCCCAGCTGAGGAGCGTC 844
    :||| :||| :||| :||| :||| :|||
146 IleLysPheGluLeuGluThrHisPheSerTyTyPrProValSerGI 162
    :||| :||| :||| :||| :||| :|||
845 TCCCGCTTCTGAAGAGGAGCAACGGGCTTCCAGCTGGCGCTGTGGCCGG 894
    :||| :||| :||| :||| :||| :|||
162 yPheValAlaProHisGlnTyLeuSerLeuGlnAspArgTyPrPheP 179
    :||| :||| :||| :||| :||| :|||
895 CTTGCTGTCCGCCCGGGAGCTTCTGCCAGAGCTGGCCTTCCGCGCTGTTC 944
    :||| :||| :||| :||| :||| :|||
179 rolLeAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
    :||| :||| :||| :||| :||| :|||
945 AGTGCACCCAGTATATCCGCCACGCGTCTCGGCCATCCACTCCCGCTGAG 994
    :||| :||| :||| :||| :||| :|||
196 ProAspLeuIleHisAspLeuGluGlyHisValProTrpLeuLeuHisPr 212
    :||| :||| :||| :||| :||| :|||
995 CCGSACTGTCTGCCACAGAGCTGTCTGGGACAGCTGCCATGCTGGCCGACCG 1044
    :||| :||| :||| :||| :||| :|||
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
    :||| :||| :||| :||| :||| :|||
1045 CACCTTGGCGCAGTTCTCCGACGAGCATTTGGC.....C 1076
    :||| :||| :||| :||| :||| :|||
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
    :||| :||| :||| :||| :||| :|||
1077 TGGCGTCCCTGGGGGCTCGGATGAGAAATTGAGAGCTGTCCACCGCT 1126
    :||| :||| :||| :||| :||| :|||
246 GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262
    :||| :||| :||| :||| :||| :|||
1127 TCA.....TGGTTTCACGGTGGAGTT 1146
    :||| :||| :||| :||| :||| :|||
262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrglyAlaValLeuI 279
    :||| :||| :||| :||| :||| :|||
1147 CGGCGTGTGTAAAGACAGACGGGAGGTGAAGGCTCATGTGTCGGGGTGC 1196
    :||| :||| :||| :||| :||| :|||
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
    :||| :||| :||| :||| :||| :|||
1197 TGTCTCTCTACGGGGAGCTCTCTGCACTGCCTGTCTGAGGAGCTTGAGATT 1246
    :||| :||| :||| :||| :||| :|||
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
    :||| :||| :||| :||| :||| :|||
1247 CGGCGCTTCCAGCCTTGAGGCTCGCGCGCTGAGCGCTTACCAAGACCAGAC 1296
    :||| :||| :||| :||| :||| :|||
312 rProGlnGluThrLeuPhePheSerIleArgHisPhe.....AspG 325
    :||| :||| :||| :||| :||| :|||
297 GTACCAGTCACTCACTTCTGCTGTGAGAGCTTCAAGTCAAGCAAGGACA 1346
    :||| :||| :||| :||| :||| :|||
325 luLeuValGluLeuThrSerLysLeuGlu 334
    :||| :||| :||| :||| :||| :|||
1347 AGCTCAGGAGCTATGCTCAGCATCCAG 1375
    :||| :||| :||| :||| :||| :|||
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-417-9
seq_documentation_block:
; Sequence 9, application US/08843417
; Patent No. 6184349
;
; GENERAL INFORMATION:
;
; APPLICANT: Herman, Ronald C
; APPLICANT: Delgado, Stephen G
; APPLICANT: Fish, Linda M
; APPLICANT: Sangameswaran, Lakshmi
; APPLICANT: Rabert, Douglas K
;
; TITLE OF INVENTION: CLONED PERIPHERAL NERVE
; TETRODOTOXIN-RESISTANT SODIUM CHANNEL
;
; NUMBER OF SEQUENCES: 10
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 525 University Ave
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U. S. A.
; ZIP: 94301

```

```
160 ValSerGlyPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAsp 176
:::|||||:::  :::::
3602 ATCTCGAACACAAAGATAAGGTGAAGACCCCTGTCTAGTACTCTCCAGCA 3553
176 gtyrPhe.ProIleAlaSer.ValMetArgThrLeuAspLys..... 189
:::|||||:::  :::::
3552 AGCTTTACCGTGGGCTTCTGTCTCCAGGTAATAGTCTTCAAAGCCAGAG 3503
190 .....
3502 ATCCACTGCTGACGAGATCATGAAGATGATGAAGCTCTCAACACGAGTG 3453
198 euIleHisasp.....LeuLeuGlyHisValPro 207
:::|||||:::  :::::
3452 TCTCCAGATACGGTAGCAAGTCTTGGCAGCTGCCAGGCCACATCCCA 3403
208 trpLeuLeu.....HisProSe 213
3402 TGGACTCTGTGTGTATCCAGTTTGCACAGGAGAGTGGCGAATGCATC 3353
213 rPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValIleG 230
|||||  |||  |||  :::::
3352 CTTCTGTGAAGCAGTATC.....TGGTTCTTCCAGGTCTATCT 3315
230 LuLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeuGln 246
:::|||||:::  :::::
3314 GCCAGCTCAGGATCTTCTCTCAGGATTTCTCAGGATCTAGGAGTCCAC 3265
247 SerAsnLeuLeuAlaIleValArg.....CysPheTrpPheThrValG 261
|||  :::::
3264 CGT.....GCTGCCCTCAGAGAGGCTTGTGTCTGCTCCACTCCCTCAG 3224
261 uSerGlyLeu..... 264
3223 CAGGGGCTCAGGAACAGATCATCTTTCCAGTCTCTACCCAGGATGGA 3174
265 .....IleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeu 278
3173 GCCAGTCTCTCAGAGATGTTCCAGTGCCTGGGCTCTGCGGTGTCAGGTG 3124
279 IleSerSerProGlnGluLeu.....G 286
3123 GTCCACACCTCTCGACTTCTGTCAGCTGCTCCGCTCTCTCTCTGGGA 3074
286 yHisAlaPheIleAspAsnValArgValLeuProLeuGluLeuAspGlnI 303
3073 TCACCTCTCTGGAAGCTCTGAGCATCTTCCCAACCATCATCTCTCCAAG 3024
303 leIleArgLeuProPheAsnThrSerThrProGlnGlnThrLeuPheSer 319
:::|||||:::  :::::
3023 TCATCAAGATCAGATTTCACCTCAGCAATGGGACAGAGACCCACACAGT 2974
320 IleArgHisPheAspGluLeuValGluLeu 329
|||  |||  |||  |||
2973 ...CGGATTAGCGATGAGTCACTGTGCTC 2947
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-540-804-11

```
seq_documentation_block:
; Sequence 11, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
; NUMBER OF SEQUENCES: 39
; ADDRESSSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
```

```

; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,804
; FILING DATE: 11-OCT-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/521,872
; FILING DATE: 21-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 241..3918
; US-08-540-804-11
```

alignment_scores:
Quality: 95.00 Length: 522
Ratio: 0.448 Gaps: 25
Percent Similarity: 40.613 Percent Identity: 18.774

alignment_block:
US-09-438-185-1047 x US-08-540-804-11 ..

Align seg 1/1 to: US-08-540-804-11 from: 1 to: 4002

```
12 TyrIleLeuLysIleAlaLeuLysLeuArgGlnSerLeuSerLeuPhePh 28
|||||  |||  |||  :::::
277 TACATCCCTGGAA.....AGCTCATCTTTGATATGACAAATCACTA 317
28 eGlnAsnSerGlnSerLeuGlnArgAlaTyrSerThrProTyrSerTyr 45
:::|||||:::  :::::
318 TAACGATTCTCAACAACGTG...CGTACGTGGAAGAGCGCATTTCTATT 364
45 yArgIleIle.....LeuGlnLysGluAsnLysGlu 55
:::  :::::
365 TTTTAAACTTTTGGGAATGTGCTACTCACTAAGATGTATCATTAAGAA 414
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePheLy 72
:::  :::::
415 .....ATCTTTCATCATTTGGCTTGTAGAGTTTATAAATAAGATGGA 455
72 sAsnLeuLeuPhe.....ValHisLeuLeuSerLeuSerLys 85
:::|||||:::  :::::
456 AAACATCGAAATTTTGGCCATTATCTTTACATATTTTGATGATTTTGG 505
85 snGlnArgGluGlyCysSerThrAspMetAlaValValSerThr..... 99
|||  :::::
506 ACGACATCTGCCCAAAATGTATACAAATGCTCTCTGTCGGCTACATAACA 555
```

```
100 .....ProPhePhe..... 102
556 TCAAGTCAAAAAGAGCCCTTCTTCTGGTAACAAAAAATCACTGATATGCT 605
102 ..... 102
606 ATTGCACAAATATTATATTGTTTCCAGCAGCAAAATCAATGATAAATGAGC 655
103 ..... 103
656 AGAATACATCATCAATCATATAAGAAAAACACACAGATAAAGTTGAAT 705
110 LeuLeuSerSerArgPheSerLeu..... 117
706 ATTCTCAAAATATTATGTTTCAAAATTTTCAAGACAATC 755
118 ..... 118
756 TTTAGAGGTGTTTATATTCCACATCACTAATGGGAAATTTACAAGCCCT 805
124 rgPhePhe..... 126
806 TACTTTTGGAAATAGTCTCAACGCCGACACTAATCAAAATTTCTGATATG 855
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHisG1 143
856 AAGAAAAAATTAGAGTTAATAGTTAGTACAGAAACGAGTCAATTGAAGAATAA 905
143 nAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrProv 160
906 TTCTTCTATACGA..... 928
160 alSerGlyPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArg 176
929 TGCTGCCAGCAACCAATGACATTTCAATTAATCTATGTCACCTGTATAA 978
177 TyrPheProIleAlaSerValMet..... 184
979 CAATTTCCAAAATCTATCAATGATCAATTAATTAATGATAGATCACTAGT 1028
185 ....ArgThrLeuAspLysAspAsn..... 191
1029 CACCAAGCTACTGGAC...GATAACCTCAGCAATTCGATTGGCCCACTT 1075
192 ..PheSerLeuThrProAspLeuIleHisAspLeuGlyHisValPro 207
076 ACGTTGACCAAAATCCCTTACAATGATCAAAATTTATCAATTAATTTCTC 1125
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG1 222
1126 TGGTCCATACATCCATCAAGCAATTTGATCACTATCAATCAAT..... 1170
222 YArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239
1171 .....CAACTGGTAGCGAATTTACTATTG..... 1197
239 ysglnArgIleGlnThrLeuLeuSerAsnLeu.....IleAlaIleVal 253
1198 .....CGAATAAATTCACAGATGAGGATTTGCACGAATTCAGATAGAA 1242
254 ArgCysPheTrpPheThrValGluSerGlyLeuIleGluAsnHisGluG1 270
1243 GATGCCATTTGGTCTATGTTTCCAA...TTAGCCAAAAAATTTTCGCG 1289
270 YArgLys..... 272
1290 CCAAAAGAGGGTGGTATCATATATGATGCTTCTTTGTATCGCTGTTA 1339
273 .....AlaTyrGlyAlaVal..... 277
1340 ATATACTAATTACTATGTCATCAATTAAGTCCCTACGTATATCAGAAAG 1389
278 LeuIleSerSerPro.....GlnGluLeuGlyHisAlaPheI1 290
```

```
1190 CTAATCAGTTCCGGCCTACTTTATCTCCAAGATTCCAATGATAAGTTTGT 1439
290 eAsp.....AsnValArgValLeuProLeuGluLeuAspG 302
1440 GCATGTCAGCTGTAAATTAACTTGAAAATTTACCCGTTGATGAAAAGTC 1489
302 I1leIleArgLeuProPheAsnThrSerThrProGlnGluThrLeuPhe 318
1490 AATACATAATGTTATTGAGGAACGTTATGGAATATGACGTTAAATTTAT 1539
319 SerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeu..... 333
1540 GAAATTTTAAATTTTCGCCCACTCGTGAATATCAGACAAATCAAAAT 1589
333 ..... 333
1590 GCGAATACTCTCCAATGATATAACTAAATTTGCAACTGTCGAAAACCTCCTC 1639
334 .....GluTrpMetLeuAsp.....Gln 339
1640 TGACATTAATAATCATGTTTGCAGAAATGTTACTTATCACATTTATGTTCC 1689
340 GlyLeuLeuGluSerIleProLeuTyrAsnGlnGluLysTyrLeuSerG1 356
1690 GGTATTTTATCTAGTGT.....AACGCCACAGTGTGCTAAAAAT 1730
356 yPheGluValLeuCys 361
1731 ATTCAGATTTTGT 1746
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-218-265-11
seq_documentation_block:
; Sequence 11, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; TITLE OF INVENTION: NO. 5922585el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/218,265
; FILING DATE: 25-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

us-09-438-185-1047.rni

Mon Jul 2 08:46:43 2001

FEATURE: ;
NAME/KEY: CDS ;
LOCATION: 241..3918 ;
US-08-218-265-11

alignment_scores: Quality: 95.00 Length: 522
Ratio: 0.448 Gaps: 25
Percent Similarity: 40.613 Percent Identity: 18.774

alignment_block:

US-09-438-185-1047 x US-08-218-265-11 ..

Align seg 1/1 to: US-08-218-265-11 from: 1 to: 4002

```

12 TyrLeuLysIleAlaLeuLysLeuArgGlnSerLeuSerLeuPhePhe 28
|||||
277 TACATCTGAA.....AAGCTCATCTTTGATGATGACAAATCACTA 317
28 ecInAsnSerGlnSerLeuGlnArgAlaTyrSerThrProTyrSerTyr 45
|||||
318 TAAGGATTCACAACTG...CGTACGTGGAGAGCAGATTCTTATT 364
45 yArgIleIle.....LeuGlnLysGluAsnLysGlu 55
|||||
365 TTTTAAACATTTGGGGAATTGCTACTCACTAAGATTGATCAATAAGGAA 414
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePhe 72
|||||
415 .....ATCTTCATCATTTGGCTTGTAGAGTTTATAATAAGATGGA 455
72 sAsnLeuLeuPhe.....ValHisLeuLeuSerLeuSerLysA 85
|||||
456 AAACCTCGAATTTTGGCCATTATCTTACATATTTTGATGATTTTGG 505
85 sGlnArgGluGlyCysSerThrAspMetAlaValSerThr..... 99
|||||
506 AGCAGATCTGCCAAATTTGATACAAATGCTCTCTGTTGGCGGTACAA 555
100 .....ProPhePhe..... 102
556 TCAAGTCAAAAGAGCCCTCTCTTCTGTGTAACAAATCACTGATGCT 605
102 ..... 102
606 ATTGCAAAATATTATATTGTTTCCAGCAGCAAAATCAATGATAATG 655
103 .....AsnArgAsnLeuTyrTyrArg 109
|||||
656 AGAACTACATCATCAATGATATTAAGAAACAAACAAAGATAAAGTT 705
110 LeuLeuSerSerArgPheSerLeu..... 117
|||||
706 ATTCTCAAAATATTATCCAGTTTAATTTTGAATTTTCAAGAACATC 755
118 .....TyrLysSerTyrCysProA 124
756 TTTAGAGGTGTTTATATTATCCACATCTAAGTGGGAAATTTTACA 805
124 rgPhePhe..... 126
806 TACTTTTGAATAGTCTCAACCCGACACTAATCAAAATCTGATATG 855
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis 143
|||||
856 AAGAAAAAATTAGAGTTAATAGTTAGTACAGAACGAGCTCAATGAA 905
143 nAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrPro 160
|||||
906 TTCTTCTATACGA.....AACGTAATAA 928
160 alSerGlyPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAsp 176

```

```

929 TGTCTGCCAGCAGCAAGCAATGACCTTTCAATTAACCTATCGTCACCTGTAAA 978
177 TyrPheProIleAlaSerValMet..... 184
979 CAAATTTCCAAACTATATCATGATCAATTAATTAATGTATAGATCACTCAGTT 1028
185 ....ArgThrLeuAspLysAspAsn..... 191
1029 CACCAAGCTACTGGAC...GATAACCTCACAGAAATTCGATGGCCACTT 1075
192 ..PheSerLeuThrProAspLeuIleHisAspLeuLeuGlyHisValPro 207
1076 ACCTTGACCAAAATCCCTTACAAATGATCAATAATTAATCAATTAATCTC 1125
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG 222
1126 TGTCTCATATCATCCATCAAGCAATTTGATCCTACTATGAATCTAAT..... 1170
222 yArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239
1171 .....CAACTGGTAGCGAAATTAATTAATCTACTATTG..... 1197
239 yGlnArgIleGlnThrLeuGlnSerAsnLeu.....IleAlaIleVal 253
1198 .....CGATAAATTCACAGATGAGGATTTGCACGAATTCAGATAGAA 1242
254 ArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGluG 270
1243 GATGCCATTTGGTCATGTTGTTTCCAA...TTAGCCAAAATTTTTCGGC 1289
270 yArgLys..... 272
1290 CCAAAAGAGGGTGGTATCATATATATGATCGCTCTTTGTATCGCCTGCTTA 1339
273 .....AlaTyrGlyAlaVal..... 277
1340 ATATACTAATTAATCTATGTCATCAATTAAGTCCCTACGTATATACAGAA 1389
278 LeuIleSerSerPro.....GlnGluLeuGlyHisAlaPheI 290
1390 CTAATCAGTTCCGGCTACTTATCTCCAAGATTCCAATGATAGTTTGT 1439
290 eAsp.....AsnValArgValLeuProLeuGluLeuAspG 302
1440 GCATGTCAGCTGTTAATTAACCTGAAATTTCAACCGTTGATGAAAGTC 1489
302 InIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLeuPhe 318
1490 AATACATATGTTATGAGCAAGTTTATGGAATATGACGTTAAATTTTAT 1539
319 SerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeu..... 333
1540 GAATTTTAAATTCGACCACTCGGGAAATTCACAGAACAAATCAAAAT 1589
333 ..... 333
1590 GCGAATACCTCCAATGATATTAATTTGCACTGTCGAAACTCCTC 1639
334 .....GluTrpMetLeuAsp.....Gln 339
1640 TGAGCATTTAAATCATGTTGTCAGAAATGTTACTTATCATTATGTTTCC 1689
340 GlyLeuLeuGluSerIleProLeuTyrAsnGlnGluLysTyrLeuSerG 356
1690 GGTATTTTATCTAGTGT.....AACCGCACAGTGTGCTTAAATAAT 1730
356 yPheGluValLeuCys 361
1731 ATCAAGATTTTGT 1746

```

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-521-872-11

seq_documentation_block:
; Sequence 11, Application US/08521872
; Patent No. 6015682
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/521,872
; FILING DATE: 31-AUG-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 241..3918
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; US-08-521-872-11

alignment_scores:
Quality: 95.00 Length: 522
Ratio: 0.448 Gaps: 25
Percent Similarity: 40.613 Percent Identity: 18.774

alignment_block:

US-09-438-185-1047 x US-08-521-872-11 ..

Align seg 1/1 to: US-08-521-872-11 from: 1 to: 4002

12 TyrIleLeuLysIleAlaLeuLysLeuArgGlnSerLeuSerLeuPhePhe 28
|||||
277 TACATCCTGGAA.....AAGCTCATCTTTGATATGACAAATCACTA 317
28 eGlnAnSerGlnSerLeuGlnArgAlaTyrSerThrProTyrSerTyr 45
|||||
318 TAACGATTCTCAACAACG...CGTACGTGAAGAGGACAGATTTCTTATT 364
45 yArgIleIle.....LeuGlnLysGluAnLysGlu 55
|||||

365 TTTTAAAACTTTTGGGAATTGCTACTCACTAAGATTGATCAATAAGGAA 414
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePhe 72
::: |||||
415ATCTTTTCATCATGGCTTGTAGAGTTTATAAATAAGATGA 455
72 sasnLeuLeuPhe.....ValHisLeuLeuSerLeuSerLysA 85
|||||
456 AAATTCGAAATTTTGCCTATTTTACATATTTTGCATGATTTTGGGA 505
85 snglnargGluGlyCysSerThrAspMetAlaValSerThr..... 99
|||||
506 ACACATCTGCCAAATGATGATCAATGCTCTGTGGCTACAATAACA 555
100ProPhePhe..... 102
|||||
556 TCAAGTCAAAAAGAGCCCTTCTTCTGTTAAACAAAAATCACTGATATGCT 605
102 102
606 ATTGCACAAATATTATTTGTTTCCAGCAGCAAAATCAATGATAATGAGC 655
103AsnArgAsnLeuTptYrArg 109
|||||
656 AGAATACATCATCAATGATATAAGAAAAACAAAGATAAAGTTGAAT 705
110 LeuLeuSerSerArgPheSerLeu..... 117
|||||
706 ATTCTCAAAATATTATCCAGTTTAAATTTTGAATAATTTTCAAGACAATC 755
118TptLysSerTyrCysProA 124
|||||
756 TTAGAGGTGTTTATATTTCACATCTAATCTGGGAAATTTCAAGCCCT 805
124 rgPhePhe..... 126
806 TACTTTTGAATAGTCTCAACGCCGACATAATCAAAATTCGTATG 855
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHisG 143
|||||
856 AAGAAAAAATTAGAGTTAATTAGTTTACAGAACAGGTCATTGAAGATAA 905
143 nAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrPro 160
|||||
906 TTCTTCTATACGA.....AAGTAAATAA 928
160 alSerGlyPheValAlaProHisGlnTyrLeuSerLeuGlnAspArg 176
|||||
929 TGCTGCCAGCAACGCAAAATGACTTTCAATTAATCACTATCGTCACCTG 978
177 TyrPheProIleAlaSerValMet..... 184
|||||
979 CAATTTCCAAAACACTATCATGCAATCAATTAATTTGATAGTACTCAGTT 1028
185ArgThrLeuAspLysAspAsn..... 191
1029 CACCAAGCTACTGGAC...GATAACCTTACAGAAATTCGATGGCCACTT 1075
192 ..PheSerLeuThrProAspLeuIleHisAspLeuLeuGlyHisValPro 207
|||||
1076 ACCTTGACCAAAATCCCTTACATGTCATAAAATATTCAATTAATTC 1125
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG 222
|||||
1126 TGTTCCATACATCCATCAAGGCAATTTGATCCTATGATCAATCTAAT 1170
222 yArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239
|||||
1171CAACTGGTAGCGAAATTTACTATTG..... 1197
239 ysGlnArgIleGlnThrLeuGlnSerAsnLeu.....IleAlaIleVal 253
|||||
1198CGAATAATTCACAGATGAGGATTTCACAGAAATTCAGATAGA 1242

us-09-438-185-1047.ini

Mon Jul 2 08:46:43 2001

```
254 ArgCysPheTrpPheThrValGluSerGlyLeuIleGluAsnHisGluGl 270
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1243 GATGCCATTGGTCATGTTGCCAA...TTAGCCAAAATTTTTCGGC 1289
      :::::::::::
270 ValGlys..... 272
1290 CCAAGAGGGTGGTATCATATGATGCCTTCTTTGTATCGCTGCTTA 1339
      :::::::::::
273 .....AlatyrGlyAlaVal..... 277
1340 ATATACTAATTACTTATGTCATTAAGTCCCTACGTATATCAGAAAG 1389
278 LeuIleSerSerPro.....GlnGluLeuGlyHisAlaPheI 290
1390 CTAATCAGTTCGGCGCTACTTATCTCCAGATCCCAATGATAAGTTGT 1439
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
290 eASP.....AsnValArgValLeuProLeuGluLeuASP 302
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1440 GCATGCCAGCTGTTAATTAATCTGAAATTTACCGTGTGCAAAAGTC 1489
302 InIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLeuPhe 318
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1490 AATACAATATGATGTTAGAGAACGTTATGGAATATGACGTTAAATTTAT 1539
319 SerIleArgHisPheASPLeuValGluLeuThrSerLysLeu..... 333
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1540 GAAATTTTAAATTCGACCAACTCGTGGAAATCAGAGAACAAATCAAAAT 1589
333 ..... 333
1590 GCGATACTCTCCAATGATATACTAATTTGCAACTGTGAAAACCTCCTC 1639
334 .....GluTrpMetLeuASP.....Gln 339
1640 TGACGATTAATCAATCAAGTTGTCAGATGTTACTATCACATTTTATGTC 1689
340 GlyLeuLeuGluSerIleProLeuThrAsnGlnGluLysThrLeuSerGl 356
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1690 GGTATTTTATCTAGTGT.....AACCCACAGCTGTTGCTTAAAT 1730
356 yPheGluValLeuCys 361
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1731 ATTCAGATTTTGT 1746
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-590-399-11
```

```
seq_documentation_block:
; Sequence 11, Application US/08590399
; Patent No. 6214588
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chaou, David M.
; TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/590,399
; FILING DATE: 26-JAN-1996
```

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
FILING APPLICATION NUMBER: US 08/540,804
FILING DATE: 11-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/521,872
FILING DATE: 31-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,265
FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03A3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 4002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 241..3918
US-08-590-399-11

alignment_scores:
Quality: 95.00 Length: 522
Ratio: 0.448 Gaps: 25
Percent Similarity: 40.613 Percent Identity: 18.774

alignment_block:
US-09-438-185-1047 x US-08-590-399-11 ..
Align seg 1/1 to: US-08-590-399-11 from: 1 to: 4002
12 TyrIleLeuLysIleAlaLeuLysLysLeuArgGlnSerLeuSerLeuPhePh 28
||||| ::| ::| ::| ::| ::| ::| ::| ::|
277 TACATCTCTGAA.....AAGCTCATCTTTGATGATGACAAATCACA 317
||||| ::| ::| ::| ::| ::| ::| ::| ::|
28 eGlnAsnSerGlnSerLeuGlnArgAlaThrSerThrProTyrSerTyrT 45
||||| ::| ::| ::| ::| ::| ::| ::| ::|
318 TACGATTCCTCAAACTG...CGTACGTGGAAGAGGACAGATTTCTTATT 364
||||| ::| ::| ::| ::| ::| ::| ::| ::|
45 yArgIleIle.....LeuGlnLysGluAsnLysGlu 55
||||| ::| ::| ::| ::| ::| ::| ::| ::|
365 TTTTAAACCTTTGGGGAATTGCTACTCCTAAGATTGATCAATAAGGAA 414
||||| ::| ::| ::| ::| ::| ::| ::| ::|
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePheLy 72
||||| ::| ::| ::| ::| ::| ::| ::| ::|
415 .....ATCTTCATCATCTGGCTTGTAGAGTTTATAAATAAGATGGA 455
||||| ::| ::| ::| ::| ::| ::| ::| ::|
72 AsnLeuLeuPhe.....ValHisLeuLeuSerLeuSerLysA 85
||||| ::| ::| ::| ::| ::| ::| ::| ::|
456 AAACCTCGAATTTTGGCATTATCTTTACATATTTTGATGATTTTGG 505
||||| ::| ::| ::| ::| ::| ::| ::| ::|
85 snGlnArgGluGlyCysSerThrAspMetAlaValSerThr..... 99
||||| ::| ::| ::| ::| ::| ::| ::| ::|
506 ACGACATCTGCCAAATGATACAAATGCTCCTGTTGGCGCTACAATAACA 555
||||| ::| ::| ::| ::| ::| ::| ::| ::|
100 .....Prophe..... 102
||||| ::| ::| ::| ::| ::| ::| ::| ::|
556 TCAAGTCAAAAAGAGCCCTTCTTTCTGGTAACAAAATCACTGATATGCT 605
||||| ::| ::| ::| ::| ::| ::| ::| ::|
102 ..... 102
606 ATTGCACAAATATTATATTGTTCCAGACGAAATCAATGATAAATGACG 655
||||| ::| ::| ::| ::| ::| ::| ::| ::|
103 .....AsnArgAsnLeuTriPTyArg 109
||||| ::| ::| ::| ::| ::| ::| ::| ::|
```


Align seg 1/1 to reverse of: 58-559-260-1 from: 1 to: 2522

35 GlnArgAlaTyrSerThrProTyrSerTyrTyrArgIleLeuGlnLys 51
||||| ||||| ||| :::: ::||| |||||
2247 GAACGTAGGTGTTCTTCATCCAGAAGCCGCTTCTGTATCTTCTTCAAAA 2198
51 sglu..... 52
1:::
2197 ACAAAAGACATCCCGAAGTCATCCGGGATGCTCTTTCTGTACATCAGACC 2148
53AsnLysGluLysGln...AlaLeuAlaArgHisLysCys..... 64
2147 GATCCAAACGAAAAAAGGAGTGTTCGCTATCTCGACACGAAGTTTCAGAG 2098
65 IleSerIleLeuGluPhePheLysAsnLeuLeu..... 75
||| ::||| ::||| ::||| ::||| ::|||
2097 ATCTTCTCCGTCAGTTATCACATCTCTCTTCTGAAATTTCAACCTG 2048
76PheValHisLeuLeuSerLeuSerLysAsnGlnArg.... 87
2047 TTTTCCGGTGTGATACGCTATGATCAGTACAGTATCCAGCCGCAATCAG 1998
88GluGlyCysSerThrAspMetAlaValVal 97
1997 TGCAGAGAAGTAATCTTCGAAGATTCAAGAACTTACACCTCTGTT... 1951
98 SerThrProPheAsnArgAsnLeuTrpTyrArgLeuLeuSerSerArg 114
::||| ||||| ::||| ||| ::||| ::|||
1950TCACCTCAGTTCCGAAGTGGTCTTCACTATGATTAAGTCTTGTATGAGGC 1904
114 gPhe.....SerLeuTrpLysSerTyrCysProArgPheP 126
1903 TTTTCCAGGTATCTTGGCTTTCCAGATCGAAGTTCCTAT.....CGTTTCT 1860
126 heLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis 142
1859 TTGCTATTATATA..... 1846
143 GlnAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrPr 159
::||| ::||| ::||| ::||| ::||| ::|||
1845CTCTTCAAGTTTTCAAAATC.....CACTTACCTTCGACGA 1808
159 ovalSerGlyPheValAlaPro..... 166
||| ::|||
1807 TGTCCTTCCATGCCCTCCCTGAAACTCTGTACTTTCTTCGATGAAT 1758
167HisGlnTyrLeuSerLeuLeuGln..... 174
757 TTGTCCAGAACCATCTTCACGAGTTGTATGCCACTTGTGAACCCGAG 1708
175AspArgTyrPheProIleAlaSerValMetArgTh 186
1707 TGCAAAGGTCTCATTCCTCGCTATGTGCCCTATGAAGAGGTCTC..... 1663
186 rLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeu 203
1662CACTTTGTAGAAGCCCTCTT..... 1642
203 euGlyHisValProTrpLeuLeuHisProSerPheSerGlu..... 216
1641CACTTCCGCATCGAAGTTGAGCCCAACCTTTGTGAAGCCTCCCGC 1597
217PhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLys 231
1596 TTTTATCACTTCGTACATTCGAAGGGTTGATCGTAGACGTTTGTGGGA 1547
231 sValGlnAlaLeuProSerLysLysGln..... 240
1546 ACTGATCGGTCTCCATCCCAAGAAGAAGGTCTCCCTGGTGTTCATCGATG 1497
240 240

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB_seq:US-08-766-014-23

seq_documentation_block:
; Sequence 23, Application US/08766014
; Patent No. 5744312
; GENERAL INFORMATION:
; APPLICANT: Mamone, Joseph A.
; APPLICANT: Davis, Maria
; APPLICANT: Sha, Dan
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766.014
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 60/008,688
; FILING DATE: December 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 223/104
; TELECOMMUNICATION INFORMATION:

```

[illegible]

11132 GTTCCCTCAAGTAATTCGGCAGGACCAAGACACTGATCATCCGAGAA 11181
315 315
11182 TGGAGCTTACGACAGCTCAGTGCATTTATCAGCACTGATCTCAAGAAG 11231
316 316
11232 TACTGCCTTAATTCGAGATATGAGACCATCAGCTGTTGTCACAGAGG.. 11279
322 sphaSpGluLeuValGluLeuThrSerLysLeuGluTrpMet..... 336
11280 .CTAATGAGATTTTACGATGTCCTCATTTTCCAGTGGTGCATTAAGA 11328
337 ..LeuAspGlnGlyLeu..... 342
11329 GGGTTGAGACCTCTGCTCTGATGTAGTGACCCCTCATGCCCCCGGAC 11378
343GluSerIleProLeuTy.....AsnGlnGlu..... 351
11379 CTGAGCGCCCATATCCCGTTATATAAGTCCCAATGATCAAAATCTTCAT 11428
352 .LysTy.....LeuSerGlyPheGluValLeuCysGln 362
11429 TAAGTACCTATGGAGGTATAGAGGATTTGTCAG 11465

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-905-817-1

seq documentation_block:
: Sequence 1, Application US/08905817
: Patent No. 582477
: GENERAL INFORMATION:
: APPLICANT: SASAKI, Keiko
: APPLICANT: MORT, Takayuki
: APPLICANT: MAKINO, Satoshi
: TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE.
: TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
: TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: YOUNG & THOMPSON
: STREET: 745 South 23rd Street
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,817
: FILING DATE: 04-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,891
: FILING DATE: 25-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/848,400
: FILING DATE: 10-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 3-293625
: FILING DATE: 14-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: PATCH, Andrew J.
: REGISTRATION NUMBER: 32,925
: REFERENCE/DOCKET NUMBER: KP-7501A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-521-2297
: TELEFAX: 703-685-0573
: TELEX: 248425 EMBON
: INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 15894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 108..1682
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..3327
FEATURE:
NAME/KEY: CDS
LOCATION: 3438..4442
FEATURE:
NAME/KEY: CDS
LOCATION: 5458..7107
FEATURE:
NAME/KEY: CDS
LOCATION: 7271..9121
FEATURE:
NAME/KEY: CDS
LOCATION: 9234..15782
US-08-905-817-1
alignment_scores:
Quality: 91.00 Length: 379
Ratio: 0.587 Gaps: 18
Percent Similarity: 40.897 Percent Identity: 20.580
alignment_block:
US-09-438-185-1047 x US-08-905-817-1 ..
Align seg 1/1 to: US-08-905-817-1 from: 1 to: 15894
57 GlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePheLysAs 73
10518 GAAGGGTTAACACATGAGCAGTGGTTGATAACTGGAAATCTTTTGTGG 10567
73 nLeuonPheValHisLeuLeuSerLeuSerLysAsnGlnArgGluGlyC 90
10568 AGTGAATT.....GGCT 10581
90 ySerThrAspMetAlaValSerThrProPhePheAsnArgAsnLeu 106
10582 GCTTTATGCTCTTAGCTGGATAGTAT.....CTGACATGTACCTA 10625
107 TrpTyArgLeuLeuSerSerArgPheSerLeuTrpLysSerTyCysPr 123
10626 AAGGACAGGCGCTTGGCTCTCCAAAGGAATGGGATTCAGTTACCC 10675
123 oArgPhePheLeuAspTyTrpLeuGluAlaPheGlyLeuLeuSerAspPheL 140
10676 GAAAGAGTTCTCTGCTTACGACCTCCCAAGGAGACCGGGTCA..... 10718
140 euAspHisGlnAlaValIleLysPhePheGluLeuGluThrHisPheSer 156
10719CGGAGGCTTGTAGATGTTTCTTAAATGATTCGAGCTTGC 10760
157 TyTyTrpProValSerGlyPheValAlaProHisGlnTyTrpLeuSerLeu 173
10761 CCATATGATGTATATGATGTTGTTAGTGGAGCTTAC.....CT 10801
173 uGlnAspArgTyTrpPheProIleAlaSerValMetArgThrLeuAspLysA 190
10802 CCAAGACCTCGATTCACCTG..... 10823
190 spAsnPheSerLeuThrProAspLeuIleHisAspLeuGlyHisVal 206
10824 ..TCTTACGCTTGAAGAAAGGAGGATCAAGGAA..... 10856


```
119 sSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuL 136
1952 ATTATAT .....CATAATTTGCTTTATTGGCAAC 1980
136 euSerAspPheLeuAspHisGlnAlaValIleLysPheGluLeuGlu 152
1981 TAGACTCTATAGATCGTTGGTCAGCATGTAAGAGATATAAGGATCG 2030
153 ThrHisPheSerTyrProVal ..... 160
2031 CTTTCATTTTGA .....CCCATGATACAATTTTCAAGAGCTATTAT 2074
161 .....SerGlyPheValAlaProHis .....GlnTyrL 170
2075 TTTTGACACAGAGAGTGTACTTACCAGTCGATTTCCCTCTTACA 2124
170 euSerLeuGlnAspArgTyrPheProIleAlaSerValMetArgThr 186
2125 AGTCAATATCGAAGATTAATCTTCTGAGATTTGGATAAAATAAT 2224
187 LeuAspLysAspAsnPhe...SerLeuThrProAspLeu.....Il 199
2175 CTTGATAAGAAAATTAATGATATCTTTCTGAGATTTGGATAAAATAAT 2224
199 eHisAspLeuGlyHisValProTrpLeuLeuHisPro..... 212
2225 GGTCTCGGTACTGGTCTGAGCTATTACGCTTTATCGTGAGCGGAATATGA 2274
213 .....SerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLys 227
2275 CTTACACATTTACGATTTCTACATTTCTGAGTCTGAGAGAAACATTACCAAAA 2324
228 .....ValIleGluLysValGlnAlaLeuProSerLysLysGlnArgIl 242
2325 GAGGAATTAATTAATTTAAGAAAGATCCCTCCACACCAACTCTT 2374
242 eGlnThrLeuGlnSer.....AsnLeuIleAlaI 252
2375 AGAATAGACAGAAACACCGGAGCATTTGACAAAGTAGCATTATTTAT 2424
252 leValArgCysPheThrPheThrValLysSerGlyLeuIleGluAsnHis 268
2425 TCATGCAACCAATCTTCCCTTTGAAACATGGGTCTCATTAAAG...TTT 2471
269 GluGlyArgLysAlaTyrGlyAlaVal 277
2472 CAAGCACCAGAGTTAGCATCTGGTA 2498
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-484-106-5

```
documentation_block:
sequence 5, Application US/08484106
Patent No. 5614618
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA #08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2700 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-484-106-5

alignment_scores:
Quality: 90.50 Length: 309
Ratio: 0.584 Gaps: 16
Percent Similarity: 50.162 Percent Identity: 20.712

alignment_block:
US-09-438-185-1047 x US-08-484-106-5 ..
Align seg 1/1 to: US-08-484-106-5 from: 1 to: 2700

24 LeuSerLeuPhePheGlnAsnSerGlnSerLeuGlnArgAlaTyrSerTh 40
1650 ATGTCGTACTTTTCCAGAAATGCTTTTCAGTA.....TTTCATTGA 1690
40 rProTyrSer.....TyrTyrArgIleIleLeuG 50
1691 CCCTCTAATCTGATTTTGAACGACGACTACTTAAATACTG... 1736
50 LysGluAsnLysGluLysGlnAlaLeuAlaArgHisLysCysIleSer 66
1737 .....AGCAGATGTCCTACA 1751
67 IleLeuGluPhePheLysAsnLeuLeuPheValHis..... 78
1752 TTCATGTTCTTTCGAGAGGTCTTATAACGACGATGCTCTGCTGACGA 1801
79 LeuLeuSerLeuSerLysAsnGlnArgGluGlyCysSer.....T 92
1802 AATCTCTTCATATTGCAACAAACAAACAGAGCCCTAGAGAGTTTGTG 1851
92 hrAspMetAlaValValSerThrPro.....PhePhe 102
1852 TTGAGTTTGTGTAAGAGAGAACCCGATTAAACGGCGATGCTAAGTTGT 1901
103 AsnArgAsnLeuTyrTyrArgLeuLeuSerSerArgPheSerLeuTrpL 119
1902 GCTCGATTCTCTCGAAGAAGAAATTGAATATAACCAATTTTAACTCTG 1951
119 sSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeu 136
1952 ATTATAT .....CATAATTTGCTTATTGGCAAC 1980
136 euSerAspPheLeuAspHisGlnAlaValIleLysPheGluLeuGlu 152
1981 TAGACTCTATAGATCGTTGGTCAGCATGTAAGAGATATAAGGATCGG 2030
```



```

31491 TCATGCAAGCAATCTTCGCTTTGTAAGACATGGGTCTCATTAAG...TTT 31537
      ::::: |||||:::
269  GUGUlyArGlySAATyTcGlyAlaVal 277
      ::::: |||||:::
31538 CAAGCACCCAGAGATTACGATCTGGTA 31564
      ::::: |||||:::

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq US-09-434-774-9

seq_documentation_block:
  Sequence 9, Application block US/09434774A
  Patent No. 6235514
  GENERAL INFORMATION:
    APPLICANT: Croteau, Rodney B
    TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPENTENYL
    TITLE OF INVENTION: MONOPHOSPHATE KINASE, AND METHODS OF USE
    FILE REFERENCE: wsr14448
    CURRENT APPLICATION NUMBER: US/09/434,774A
    CURRENT FILING DATE: 1999-11-04
    NUMBER OF SEQ ID NOS: 16
    SOFTWARE: PatentIn Ver. 2.0
    SEQ ID NO 9
      LENGTH: 1203
      TYPE: DNA
      ORGANISM: Lycopersicon esculentum
      FEATURE:
        NAME/KEY: CDS
        LOCATION: (1)..(1203)
US-09-434-774-9

alignment_scores:
  Quality: 89.50      Length: 349
  Ratio: 0.963      Gaps: 19
  Percent Similarity: 45.559      Percent Identity: 22.923

alignment_block:
US-09-438-185-1047 x US-09-434-774-9  ..

Align seg 1/1 to: US-09-434-774-9 from: 1 to: 1203

68  LeuGluPhePheTyrLysAsnLeuLeuPheValHisLeuLeuSerLeuSerIly 84
43  TTGATTCTTTTAAAGAGATGTTTTTTTCAGTCATGGTCA.....84
84  sAsnGlnArgGluGlyCysSerThrAspMetAlaValSerThrProp 101
85  .AATAGGCCACATGGTTCATCC.....T 106
101  hePheAsnArgAsnLeuTyrTrpArgLeuLeuSer.....112
107  ATTTTAAACAAGAATATCAATTTTAGAAGAAACAGTTTGTATTGTGAAG 156
113  .....SerArgPheSerLeuTyrLysSerTyrCysPr 123
157  GCTTCAGGTTCAAGAACTAGTAAAAACAAGTAGAGATAACATATAATCC 206
123  oArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheI 140
207  TGA.....GAGAGTTTAAATAATGCTGATGAAG 238
140  euAspHisGlnAlaValIleLysPhePheGluLeuGluThrHisPheSer 156
239  TGGATAGAGAAGCTGGGTTTTCAAGACTCACCTCT.....TTTCT 279
157  TyrTyrProValSerGlyPheValAlaProHisGlnTyrLeuSerLeuLe 173
280  CCTTGCAGATAAATGTTTCTTG.....AGATAACAAGCAAGAG 320
173  uGlnAspArgTyrPheProIleAlaSerValMetArgThrLeu.....187
321  GGATGCGGATATCATGATTGGCGGTCTCTCTTCATGATTAATAGTCTAG 370

```


OM of: US-09-438-185-1047 to: GenEmbl.* out_format : pfs
 Date: Jun 29, 2001 2:47 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framel_p2n.model -DEV=xlp
 -Q/cgn2.1/usptc_spool/US09438185/runat_29062001_074544_6960/app_query.fasta_1.430
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.500 -FGAPOP=6.000
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=200000000 -USER=US09438185 -CGN1_1_5400 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPX -WAIT -THREADS=1

Search information block:

Query: US-09-438-185-1047

Query length: 362

Database: GenEmbl.*

Database sequences: 1344157

Database length: -85606004

Search time (sec): 1336.620000

score_list:

Sequence	Strd Orig	zscore	Escore	Len	Documentation
gb_ba1:AE001685	+ 1889.00	2691.87	1.1e-141	10847	AE001685 Chlamydia pneumoniae
gb_ba1:AE002240	+ 1889.00	2684.54	2.7e-141	22589	AE002240 Chlamydia pneumoniae
gb_ba1:AE002548	+ 1889.00	2657.85	8.3e-140	325865	AE002548 Chlamydia pneumoniae
gb_ba1:AE004410	+ 256.50	341.08	9.2e-11	10143	AE004410 Vibrio cholerae
gb_ov:GGA251387	+ 237.50	331.97	3.0e-10	1633	AJ251387 Gallus gallus mRNA for
gb_ov:RRTPH2	+ 236.50	321.88	1.1e-09	3877	X53501 Rat mRNA for tryptophan
gb_ba1:AE004522	+ 236.50	307.88	6.5e-09	15724	AE004522 Pseudomonas aeruginosa
gb_ba1:HSWHYDR	+ 231.50	325.34	9.4e-10	1335	X52836 Human mRNA for tryptophan
gb_ov:GGU26428	+ 231.50	322.96	9.4e-10	1757	U26428 Gallus gallus tryptophan
gb_ov:MUSTPH	+ 230.50	321.29	1.2e-09	1732	J04758 Mouse tryptophan hydrox
gb_ov:PEPPHY	+ 230.50	316.34	2.2e-09	2843	M88627 Pseudomonas aeruginosa
gb_ov:RATTOHAB	+ 229.50	322.46	1.0e-09	1335	L29305 Homo sapiens tryptophan
gb_ov:RATTOHAB	+ 229.50	321.90	1.1e-09	1412	L29305 Oryctolagus cuniculus t
gb_ov:RATTOHAB	+ 229.50	318.04	1.1e-09	2077	M2478 Quail tyrosine hydroxyl
gb_ov:RATTOHAB	+ 229.50	318.04	1.1e-09	1757	M2478 Quail tyrosine hydroxyl
gb_ov:RATTOHAB	+ 229.50	313.95	3.0e-09	1757	M2478 Quail tyrosine hydroxyl
gb_ov:RATTOHAB	+ 229.50	310.73	4.5e-09	2425	AJ000731 Anguilla anguilla mRN
gb_ov:RATTOHAB	+ 229.50	312.01	3.8e-09	1847	AF031034 Schistosoma mansoni t
gb_ov:RATTOHAB	+ 229.50	311.69	4.0e-09	1651	L20679 Xenopus laevis tryptoph
gb_ov:RATTOHAB	+ 229.50	310.27	4.8e-09	1770	M10244 Rat tyrosine hydroxylas
gb_ov:RATTOHAB	+ 229.50	308.51	6.0e-09	2111	L22651 Rat tyrosine hydroxylas
gb_ov:RATTOHAB	+ 229.50	308.54	6.0e-09	1696	M36794 Bovine tyrosine hydroxyl
gb_ov:RATTOHAB	+ 221.50	304.90	9.6e-09	2441	M17250 Rabbit tryptophan hydro
gb_ov:RATTOHAB	+ 221.50	305.05	5.7e-06	332635	AP030305 Mesorhizobium loti
gb_ov:RATTOHAB	+ 220.00	305.70	8.6e-09	1816	X05230 Human mRNA for tyrosine
gb_ov:RATTOHAB	+ 220.00	305.29	9.1e-09	1891	Y00414 Human mRNA for tyrosine
gb_ov:RATTOHAB	+ 220.00	305.14	9.3e-09	1921	M17589 Human tyrosine hydroxyl
gb_ov:RATTOHAB	+ 219.00	305.21	9.2e-09	1650	X76209 D.melanogaster mRNA for
gb_ov:RATTOHAB	+ 216.50	302.61	1.3e-08	1494	AF030336 Schistosoma mansoni t
gb_ov:RATTOHAB	+ 214.00	301.51	1.5e-08	1193	K02599 Rat phenylalanine hydro
gb_ov:RATTOHAB	+ 213.50	296.10	3.0e-08	2429	K03020 Human phenylalanine hyd
gb_ov:RATTOHAB	+ 213.50	293.43	4.2e-08	2429	K03020 Human phenylalanine hyd
gb_ov:RATTOHAB	+ 213.50	292.45	4.7e-08	2679	L47726 Homo sapiens phenylalan
gb_ov:RATTOHAB	+ 212.50	294.14	3.8e-08	1959	X51942 Mouse mRNA for phenylal
gb_ov:RATTOHAB	+ 212.50	294.08	3.8e-08	1706	M36705 Bovine tyrosine hydroxyl
gb_ov:RATTOHAB	+ 211.50	294.08	3.8e-08	1706	M36705 Bovine tyrosine hydroxyl
gb_ov:RATTOHAB	+ 210.00	272.19	6.3e-07	12277	AF340166 Streptomyces verticil
gb_ov:RATTOHAB	+ 207.00	288.69	7.6e-08	1529	M32802 D.melanogaster phenylal
gb_ov:RATTOHAB	+ 206.50	287.26	9.2e-08	1642	AF135186 Caenorhabditis elegans
gb_ov:RATTOHAB	+ 197.00	274.82	4.5e-07	1449	AF18333 Drosophila melanogaster
gb_ov:RATTOHAB	+ 196.50	278.93	2.7e-07	894	AF146711 Chromobacterium violace
gb_ov:RATTOHAB	+ 193.00	235.80	6.8e-05	40392	U21308 Caenorhabditis elegans

gb_ba1:AE005836 - 189.50 242.72 2.3e-05 12203 AE005836 Caulobacter cresc
 gb_ba2:CHRPHEHYD + 187.00 262.25 2.3e-06 1207 M55915 Chromobacterium viol
 gb_in3:DMPEHY + 182.00 243.62 2.5e-05 3786 X98116 D.melanogaster pah/t
 gb_in3:AC019368 + 182.00 210.64 0.0017 102552 AC019368 Drosophila melan

seq_name: gb_ba1:AE001685

seq_documentation_block:

LOCUS AE001685 10847 bp DNA BCT 01-DEC-2000

DEFINITION Chlamydia pneumoniae section 101 of 103 of the complete genome.

ACCESSION AE001685 AE001363

VERSION AE001685.1 GI:4377378

KEYWORDS

SOURCE Chlamydia pneumoniae CWL029.

ORGANISM Chlamydia; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 10847)

AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,

Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.

Comparative genomes of Chlamydia pneumoniae and C. trachomatis

Nat. Genet. 21 (4), 385-389 (1999)

TITLE

JOURNAL

MEDLINE

PUBMED 99206606

REFERENCE 2 (bases 1 to 10847)

AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,

Grimwood,J., Davis,R.W. and Stephens,R.S.

Direct Submission

Submitted (01-DEC-1998) Program in Infectious Diseases, University

of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

Location/Qualifiers

1. .10847

/organism="Chlamydia pneumoniae"

/strain="CWL029"

/db_xref="taxon:115713"

complement(134) .826

/gene="Cpnl045"

complement(134) .826

/gene="Cpnl045"

/codon_start=1

/transl_table=1

/product="conserved hypothetical bacterial membrane

protein"

/protein_id="AADI9182.1"

/db_xref="GI:4377378"

/translation="MNTSHKTLVFLVSTLTLVLSNEVLSKLTPTTFNFIP

GGILYPTLFTISDVVNEIFGPKKARVIFSAFANLASSIVQIFMFPVSPMOT

AMHCLFDLPRLFLASLLAFIVSQDLIVLYTFKRNTPNSLWRSNGSTWISQIPD

TFIVTQCLTYRGMGLSPQTLNIMFYIYKTFVCLVLTPLFLAVNTIRKFLGMPST

KIANVPLINOP"

complement(823) .1911

/gene="Cpnl046"

complement(823) .1911

/gene="Cpnl046"

/codon_start=1

/transl_table=1

/product="Aromatic Amino Acid Hydroxylase"

/protein_id="AADI9183.1"

/db_xref="GI:4377380"

/translation="MHYCELTDPKYLKIALKQSLFLFQNSQSLQRAYSTPYSY

NRILYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

YRLLYQKNKQALARKKICILEFFKLLFVHLLSKNRCQECSTDMVAVSTPFF

```

/protein_id="AAD19184.1"
/db_xref="GI:4377381"
/translation="MGSMHGVGVGCSGRTGKIVSALEQSSYITLPGFSSSALTIL
FOVIANVDFSHPLIKTEVVAHLISPKPLIIGTGPCKKEAHDSEELTHIV
PVVVCNAGLGYIHKRLVMLSLQCNQPNQIDIRITHRYKKDKSLSGTAQDLDTIQ
VQGEQWGEYEVGQDSKKTIEVQSSRVGDIPEGHEVAFISSGQILVRHTVFSRN
VQGRGLSILDWKLTLNPPQGLYSLGDTLELVLNRHECLLKKTTDH"
/gene="asd"
/notes="3843"
/gene="cpn1048"
/notes="3843"
/gene="asd"
/codon_start=1
/transl_table=11
/product="Aspartate Dehydrogenase"
/protein_id="AAD19185.1"
/db_xref="GI:4377382"
/translation="MRIAVALGVTLGVQKVFALLHKWRDVIAEVVASNKYGQSYG
DACIMQEPGPMPWRDLPINKIEEVOSDIIVSFLPSSAESMEAYCLSQGKVFPSNA
STYRMSSVPIIIPVNSDHFOLLEEQPIPKIITSNCCVSGITLALAPLRKFLSDH
VHVTLOSAGGAGYGVFSLDLLANTVPHVIGEEKILRETVKILGSSKQPLPCKLSV
TVHRVPVAYHTLSLVTFKSDVDLDEILYSQEKNEFPNTYQILDNPWPSPQARKHL
SHDDMRVHLGPTIGDGFRTKMNVLIHNLVRGAAGTLLASMENYFFDYLRKEMCLR"
/3831..5153
/gene="lysc"
/notes="Cpn1049"
/3831..5153
/gene="lysc"
/codon_start=1
/transl_table=11
/product="Aspartokinase III"
/protein_id="AAD19186.1"
/db_xref="GI:4377383"
/translation="MSKIVYFGTSLATFNICLVCDIICKDKSFVVVSAIAGVTD
LIVDCSSILREBEVLKIEGHEEVKNLAIPFPVSWTWSRLPYLQHLIEISLDF
ARLISGEDISLVRVCSGWDGLFLEARSVLHDDSYRASPNLDMKAHHQL
ELMQPSTIIQGGISNGEIVLIGGSDISATLAEARAFETVIRVDNVIITMD
PVIYDAQRIPELSEPMONLAKGAKVLPMPDFPCMRAGIPFIYFSTDEPKRGTW
YAVDSVSEPRIKALSLSQSFSDVDTVLVCGGLEIIGLDESHGIDPELMAIQ
NNVGVPMDDIIISQAEQHLVDLSLSSVTRLHLSVALITMIGDNLNSPKPVSVTITE
5162..6037
/gene="dapA"
/notes="Cpn1050"
5162..6037
/gene="dapA"
/codon_start=1
/transl_table=11
/product="Dihydrodipicolinate Synthase"
/protein_id="AAD19187.1"
/db_xref="GI:4377384"
/translation="MHLITATVPFPNGTIDFASLERLLSFODAVNGVVLGSGTGE
GLSLTKKEQALICFACDLQKLPFLVGTSTLLEEVLDWIHFCNDLPISGLMTPTPI
YTRPKLGGQILWFPAVNAAKHPALLYNIPSRATPLYLDTVKALAHPOFLGKDSG
GSVEEFQSKYSTAPHIQLYCGDDVFWSEMAACGAGHLISVLSNAPPEAREVYNPOE
QDYSLSAMETCRWVTNTPIGIKAILAYKKAITHAQLRLPLSIEDFDLENVPVAVES
MLAWPKLRVSFYSYS"
/6198..6509
/gene="Cpn1051"
/6198..6509
/gene="Cpn1051"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19188.1"
/db_xref="GI:4377385"
/translation="MTPKSIQQLHLIKTIDFVRKISPVTTKKSSFFRQSLRFLFLFW
MFLYCIIRSFCHVIATFICRGLILFTLTLFSLMICILHFTILPWLICKEDPRIIRKN
K"
/6553..7408
/gene="Cpn1052"
/6553..7408
/gene="Cpn1052"

```

```

/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19189.1"
/db_xref="GI:4377386"
/translation="MKYNSREKIKSALRICSSYICITVFRNFFSLSCYDKIFYLSLCYV
FNGNSIGCRKSCFCFRGKKTVEVTEKVIKDEIRPSELEGNDPVKVAESFPKRRRALE
SLSSQSIGNLCAISNFDLSQMLSNFNSKEIWSGTITFRSKSTCDAEGSPFPRTYACG
YLAGRLSKLAYSVELGVTAGLQGRLLKDVSDSHRTRATSSILSVHGSMTVRPLSCTKY
IVGKARELLPFFRLTSDVRRDLKKKFRLECKD"
/7389..7940
/gene="Cpn1053"
/7389..7940
/gene="Cpn1053"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19190.1"
/db_xref="GI:4377387"
/translation="MNFADHNLHLYLCLGLDGVACPILLSLDCPLNYSKASHEVLV
YSKFRCSIGBPSRLATSGNDTYYSIVSLPIGLRYEVTSPSGRHDNDIMHVAPKIGV
LSHGTREAKEIPGSSKDYAFFSLTARESLMISEKLAMTFOVSEVQNCYSQCTKVTKT
NLKEQYRHLSHNTGPFELSVKSAF"
/8270..10705
/gene="Cpn1054"
/8270..10705
/gene="Cpn1054"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19191.1"
/db_xref="GI:4377388"
/translation="MOVLLSPOLPPPOHSGVSISSPSKLRVLAINFLVFGMLLISG
ALFTLIGPGLSAAISGLGICLSALGVLMISGLILLYKREIPVRPEEIPGCVSL
APSEPALQAOKTLAQLPKELDQDLDIOEVFACILKDKSKYSESLNDAKLELR
PFDVWEDILSEIFELQIYVAQGWDLNLFNGSLMTAASESLDLFHVSKRLCYL
VSGVREGGLSKAKEIARLMSLHCEIHKVAVAFDRNSTAMAEKAFKALRALESV
YRSUTQYRDLFESERAKIPWNGHITWLRDDAKSGCAEKLRDAERMKKFRVFW
VEEDQDINLLGDGTVIDPYRQRMDEITFHELYEKTFFLRLHRCALAKTTFE
KKRKNLKAIEENYANRCVREQEDYWKEEKREAREFERNGKILSPVEELESLEQFD
TRSNLEKAYAEENYANRCVREQEDYWKEEKREAREFERNGKILSPVEELESLEQFD
HGLNFSEKLEMELEHILKQAEATAVENKILSDAESRLEIFEDVKEMPCRIEIE
KTLRMAELPLLTPKAFKACSOYNACAMLEKVKPYCKESLAYVTSKERLVSLEDDL
RRAYTECQRFQDGSLESEVACREQLREIQEFETQGLDLVEKELLCVSSLRNTE
CDCVSGVKEAPEPKKFFYAYYDIYRVVQSRWMTSERLREGVQACNMKLAGUSE
EDVKLEEEYWLRYEERKNKRLVGTIVATQORVAAFESIEVPEIPEAPEKPSLL
DKARSLFTREDHT"
/3261 a 1883 c 2436 g 3267 t
BASE COUNT
ORIGIN
alignment_scores:
Quality: 1889.00 Length: 362
Ratio: 5.218 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-438-185-1047 x AE001685/rev ..
Align seg 1/1 to reverse of: AE001685 from: 1 to: 10847
1 ValHistTyrCysGluArgThrLeuAspProLysTyrIleLeuLysIleal 17
1911 GTGCACCTACTGCGAGAGAACCTGGACCCAAAGTATATCTGAAGATTGC 1862
17 aLeuLysLeuArgGlnSerLeuSerLeuPheGlnAsnSerGlnSerL 34
1861 TCTAAAGCTGAGACATCACTTTCCTCTTCTCCAGAACAGCAATCAC 1812
34 euGlnArgAlaTyrSerThrProTyrSerTyrTyrArgIleIleLeuGln 50
1811 TCCAAACGTGCATCTCGACCCCATATCTCTACTACCGAATCATCTACAA 1762

```

51 LysGluAsnLysGluLysGlnAlaLeuAlaArgHisLysCysIleSer11 67
1761 AAGGAAATAAAGAGAGCAAGCTTTAGCTCGACAAATGCATTTCTAT 1712
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
1711 TTTAGAAATTTTCAAAACATCTCTTTGTTTCATCTTCTGCTATCA 1662
84 YAsnGlnArgGluGlyCysSerThrAspMetAlaValValSerThrPro 100
1661 AGAATCAAGAGGAGGTGCTCAGCTGATGCTGTTGTGAAGCACTCC 1612
101 PhePheAsnArgAsnLeuThrPyrArgLeuLeuSerSerArgPheSerLe 117
1611 TTTTAAATCGGAATTTATGATATGCTGCTCTTCTCAGGTTTCTCT 1562
117 uTrpLysSerThrCysProArgPhePheLeuAspThrLeuGluAlaPheG 134
1561 ATGGAAGAGCTATGTCAGAGATTTTCTGATCTACTAGAACCTTCG 1512
134 yLeuLeuSerAspPheLeuAspHisGlnAlaValLileLysPheGlu 150
1511 GTCTCTTCTGATTTCTAGACCAACAGCAGTCATTAATTTCTCGAA 1462
151 LeuGluThrHisPheSerThrPyrProValSerGlyPheValAlaProH 167
1461 TTAGAAACACATTTTCTTATATCCGTTTCAGGATTTGTAGCTCCCA 1412
167 sGlnThrLeuSerLeuLeuGlnAspArgThrPheProIleAlaSerVal 184
1411 TCAATACTGTCTGTGTGAGGACCGTTACTTTTCCCATTTGCCCTCTG 1362
184 eArgThrLeuAspLysAspAspPheSerLeuThrProAspLeuIleHis 200
1361 TCGCAACTCTCGATAAAGATAATTTCTCTTAACTCTCATCTCATCAT 1312
201 AspLeuLeuGlyHisValProThrLeuLeuHisProSerPheSerGluPh 217
1311 GACCTTTTAGGACGTCGCTTGGCTTCTACATCCCTCATTTTCTGAAT 1262
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGlnA 234
1261 TTTCAATACATGGAGAGATCTTCTACTAAATCATAGAAAGATGACAG 1212
234 laLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
1211 CTCTTCTTAGTAAAGAAACACGATACCAACCTACAAGCAATCTGATC 1162
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267
1161 GCTATTGTACGCTGCTTTGTTGTTACTGTTGTAAGACGACTATTGAAA 1112
267 nHisGluGlyArgLysAlaThrGlyAlaValLeuIleSerSerProGlnG 284
1111 CCATGAAGAGAGAAAGCATATGAGCCGCTTCTTATCAGTTCTCTCCAG 1062
284 luleGlyHisAlaPheIleAspAsnValArgValLeuProLeuGluLeu 300
1061 AACTTGCACACGCTTTCATTTGATAGACGTACGTTCTCCCTTTAGAATT 1012
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLe 317
1011 GATCAGATTTATTCGCTTCTCCCTCAATATACATCACTCCACAGAGACTT 962
317 uPheSerIleArgHisPheAspGluValGluLeuThrSerLysLeuG 334
961 ATTTCATAGACATTTTGAAGACGTTGAGAACCTCACTTCAAAATTAG 912
334 luTrpMetLeuAspGlnClyLeuLeuGluSerIlePheProLeuThrAsnGln 350
911 AATGGAGTCTGACCAAGGCTGTGTAGAAATCAATCCCTTTTACAAATCA 862
351 GluLysThrLeuSerGlyPheGluValLeuCysGln 362

861 GAGAAATATCTTCTGTTTGGAGTACTTTGCAA 826

seq_name: gb_bai:AE002240

seq_documentation_block: 22589 bp DNA BCT 30-MAY-2000
LOCUS AE002240
DEFINITION Chlamydomophila pneumoniae AR39, section 68 of 94 of the complete genome.

ACCESSION AE002241 AE002161
VERSION AE002240.2 GI:8163502

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE


```
832 GTGCACTACTGCGAGAGAACCCCTGGACCCAAAGATATATATCTGAAGATTGC 881
17 aLeuLysLeuArgGlnSerLeuSerLeuPhePheGlnAsnSerGlnSerL 34
882 TCTAAAGCTGAGCAATCACTTCCCTGCTTCTCCAGACAGCAATAC 931
34 euGlnArgAlaTyrSerThrProTyrSerTyrArgGlnLeuLeuGln 50
932 TCCAAACGTGCATCTCGACCCCATATCTCTACTACCGAATCATCTTACAA 981
51 LysGlnAsnLysGlnLysGlnAlaLeuAlaArgHisLysCysIleSerI 67
982 AAGCAAAATAAAGAGAACGCTTTAGCTCGACACAAATGCATTTCTAT 1031
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
1032 TTTAGATTTTCCAAACACTACTCTTGTTCATCTTCTGCTATATCAA 1081
84 yAsnGlnArgGluGlyCysSerThrAspMetAlaValSerThrPro 100
1082 AGAATCAAGGAGGAGGTCTCCACTGATATGGCTGTGTAGCACTCCC 1131
101 PhePheAsnArgAsnLeuTrpTyrArgLeuLeuSerSerArgPheSerLe 117
1132 TTTTAAATCGGAATTAATGGATCGACTCTTCTCAGCGTTTCTCT 1181
117 utrPlySerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
1182 ATGGAAGAGCTATTGTCCAGATTTTCTTGTATCTTGAAGACTTCCG 1231
134 lYLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPhePheGlu 150
1232 GTCCTCTTCTGATTTCTTAGACCATCAAGCAGTCATTAATCTTCGAA 1281
151 LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProHi 167
1282 TTGAAGACACATTTTCTTATATCCGTTCTCAGGATTTGTAGTCCCCA 1331
167 sGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerValM 184
1332 TCAATACTTCTGCTGTGTCAGGACGGTACTTCCCATTCGCTCTGTAA 1381
184 eArgThrLeuAspLysAspAspPheSerLeuThrProAspLeuIleHis 200
1382 TCGGAATCTCGATAAGATAATTTCTCTTACCTCCATCTCATCATCAT 1431
201 AspLeuLeuGlyHisValProTyrLeuLeuHisProSerPheSerGluPh 217
1432 GACCTTTTAGGCGACGGCTTGGCTTCTTACATCCCTCATTTCTGAAT 1481
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGlnA 234
1482 TTTTATAAACATGGGAAGACTTCTACTAAGTCTATAGAAAGATCAAG 1531
234 lAlaLeuProSerLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
1532 CTCTCTCTAGTAAACAAACAGCATACAAACCCCTACAAAGCAATCTGATC 1581
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267
1582 GCTATTGTACGCTCTTTTGGTTTACTGTGTAAGCGGACTTATTGAAA 1631
267 nHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnG 284
1632 CCATGAGGAAGAAACATATGAGCCGTTCTTATCATCTCTCTCAGG 1681
284 lLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGluLeu 300
1682 AACTTGACACCTTCTTATGATAACGTACGTGTCTCTCCCTTAGAATTC 1731
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLe 317
1732 GATCAGATTTATTCGTTCTCCCTCAATACATCACTCAACAGAGACTTT 1781
```

```
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334
1782 ATTTTCAATPAAGACATTTTGTAGTGAACCTCACTTCAAAATTAG 1831
334 luTrpMetLeuAspGlnGlyLeuLeuGluSerIleProLeuTyrAsnGln 350
1832 AATGGATCTCACCACAGTCTGTAGTAATCAATTCCTCTTACATCAA 1881
351 GluLysTyrLeuSerGlyPheGluValLeuCysGln 362
1882 GAGAAATATCTTCTGCTGTTGAGGTACTTTGCCAA 1917
seq_name: gb_ba2:AP002548
seq_documentation_block:
LOCUS AP002548 325865 bp DNA BCT 08-DEC-2000
SOURCE Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
DEFINITION section 4/4.
ACCESSION AP002548 AB033788 AB033790 AB033791 AB033818 AB036083
AB036084 AB036086 AB036087 AB036088 AB036089 AB038350
AB038351 AB038352 BA000008
AP002548.2 GI:10176694
VERSION
KEYWORDS
SOURCE Chlamydomophila pneumoniae J138 (strain:J138) DNA.
ORGANISM Chlamydomophila pneumoniae J138
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,F., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
REFERENCE 2 (bases 1 to 325865)
MEDLINE Direct Submission
AUTHORS Shirai,M.
TITLE Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
JOURNAL Mutunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
COMMENT On Sep 15, 2000 this sequence version replaced gi:6172302
gi:6172304 gi:6172306 gi:6172308 gi:6172400 gi:6635182 gi:6635184
gi:6635186 gi:6635188 gi:6635190 gi:6635192 gi:6635194 gi:8547442
gi:8547447 gi:8547451 gi:8979170.
AB033788-AB033791, AB033818: Submitted (25-Oct-1999)
AB038350-AB038352: Submitted (14-Feb-2000)
AB036083-AB036089: Submitted (18-Dec-2000).
FEATURES
source
1. 325865
/organism="Chlamydomophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/annotation="Chlamydia pneumoniae (strain J138)"
221..1318
/feature="CPj0797"
221..1318
/feature="CPj0797"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAA99005.1"
/db_xref="GI:8979171"
/translating="MSKKIKVLGHLTCTLFRGVLCAALSNIGYASTQESPYQKSI
EDWKYTFDLELLSKESGESEAHISNGSRIVGASGOGSVATVWESHILKHLGT
LGGEASSAGISKDEGVVGVWGTREGYTHAFVDFDRMDKDLGTIGATYSVARGVSG
GSIIVGVSATARGEDGWQVKGWKIKQLKLLPQLGWSANAISETDVTIVGRGE
ISRNHIVAKWNKNVATSLGTGLGSVASAEIISANGKIVGWSVNGSTNNGETHAFMHKDE
TWHDLGTGGFSVATGVSADGRAIVGFSVAVTGEIHAFFYAEEMEDTLTGGEGAR
VFDISSEGNLIIGSIKTDAERAYLFTHK"
1467..2480
/feature="CPj0798"
gene
```

CDS	1467..2480 /gene="CPj0798" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA99006.1" /db_xref="GI:8979172" /translation="MKKTCQNYRIGVWFVFLVLTOTLPAGHFDIGTSGLYSW ARGVSDGVVYGVGGNAPKVDGEGKLEGLGVPRLSEALVFKASVDGSGVIGISDOD PSCRKVVNGALVDLGFSEGMOSFAGVSSGDKTIVGCLASDPRFNFAVKWDETC MYVLNPLNDRSCANDSEDSGVIVGDANGSEIKAKYAKWKGQHLILSNLPGAKRS SAHVSXKDSGFIVGEFISEENENAVFYHNCYIKDITGLGSDYSVATGVSRCKVIYG HSFTDCEIRAKYVDGRMIDLGLGGSASFATGVSDDGKTIIVGKFEIETELGCHAFII LDD complement(2558..3607) /gene="CPj0799" /complement(2558..3607) /gene="CPj0799" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA99007.1" /db_xref="GI:8979173" /translation="MAAIKQILRSLMSQSSLMWVLSLXSGYCYVITDKPEDDFHS SSAVKWDHMKTTLSRLNKRKASAKAVSGTGATTVGFIKDTWSRTYAVRWNVGFKEL PTSGWVKKATGISDSGSIILAGIVNELSQSAFTWKNMNYLLPSTWAVOSKAYGI SSDGSVIVGSANDRTFAVKWTGHEAOVLVPGWAKVSVANSVANGSIILVGSVODA SGILYAVKWEGNTITHLGTGLGYSALAKAVSNCKKVIIVGRSETTYGEVHAFCKHGVN SDLGTGGSYSAKGVSGATGVIVGMSTANGKLHAFKYVGGRMIDLGYSWKACAN AVSDGEIIVGVQSE" complement(3867..5153) /gene="eno" /complement(3867..5153) /gene="eno" /codon_start=1 /transl_table=11 /product="enolase" /protein_id="BAA99008.1" /db_xref="GI:8979174" /translation="MFEAVIADIQAREILDSRGYPTLHVKKVITSTGSGVEARVPSGAS TGKKEALEFRDPSRYQCKGVLAQVKNVKELFPLVKGCSYVEQSLIDSLMDSGDS PNKETAGANALIGSLATAHAAATLRPLYRYLGGCFACSLPCPMNLINGMHADN GLGFQFMIRPIGASSIKREAVNMGADVFTLKLHERGLSTGVGDEGFGAPNLASNE EALELLAIEKAGFTPGKDISLADCAASSFVNKVTGYDGRHYEEQIAILSNLCDR YPIDSTEDGLADEYDGNALLTEVLGKQVIVGDLLFTVNPPLILEGISNGLANSVLI KPNQIGTLTETVATKLAQMAGYTTIISHRSGETTDTTADLAVAFNAQIKTGSLSR SERVAKYRNLMEIEELGSEALFTDSNVFSEVDEE" complement(5345..7318) /gene="uvrB" /complement(5345..7318) /gene="uvrB" /codon_start=1 /transl_table=11 /product="exonuclease ABC subunit B" /protein_id="BAA99009.1" /db_xref="GI:8979175" /translation="MTFQLHAPFACGQDQPEATARLSAGVRNOKVSQVLLGTGSGKT FTIANVAVNLTPLVAHNKTLAAQYQEFREFFPNNAVEFYISYDIYQPEAYIAR SDYIIEKSLINDEIKLRLSATRSILERRDTLIVSVSCIYIGSPENYTSMAVLVE VGEYEPRLITQALVKMHYQASPIQPSAFRERGSVIDIPPAYESELARLREFLNDTL TSEYSDPLTMIKESVPSATLYPGSHYVIEPAIRQOARTIQEELERMAFFDDRPI EKDRIEFTHTDIEMKEIGFCCKGIENYSRHFTGAPGAPPTCLLDYFDFPLIIDE SHOTLPQIRAMYGQSRKQSLVEYGFRLPSAFDNRLPLYEEAQYFRKYIYVSATPG TDVQESSGHIYVQIIRPTGIPDMPPEIRPATQVDDLLLEEIRLSQKHEKILVISI TKKLADMGAGFISELEIPAAVLSHGISQFCEAARNITQITDLRSQVIDVLGVNLRGLDL PEYSLVALIDADKEGFLRSTSLTQFCGRAARNNGKVIYFADQKTRSTIEETLRETER RRQIQDYKNEINVPKPIKAIKIANPILQTSKDSPEKSRPLSKEDLEQIKKYE AMORAKREFRENAKYRDAQACKEQLLYLF" complement(7327..8361) /gene="trpS" /complement(7327..8361) /gene="trpS" /codon_start=1	gene CDS	1889.00	Quality: 362
gene CDS	complement(8370..8924) /gene="CPj0803" /complement(8370..8924) /gene="CPj0803" /codon_start=1 /transl_table=11 /product="CT584 hypothetical protein" /protein_id="BAA99011.1" /db_xref="GI:8979177" /translation="MAAKTKLEEDNVFLLEGNLKRIFATPIGTYTTFREFONVVFN CANGQOBIANFFFEMLINKLQELAPOQKQAHSLIAEFMMPIRVAKDITHEREGEIN FITSDMLTQERCIFLNRLARVDGQEFLLMTDVQNTCHLIRHLARLEAKQNPVGEK NLQIOEITSLKNHFDLTKALQ" complement(8928..9680) /gene="gp6D" /complement(8928..9680) /gene="gp6D" /codon_start=1 /transl_table=11 /product="CHLTR plasmid paralog" /protein_id="BAA99012.1" /db_xref="GI:8979178" /translation="MGNLATLLESRFKNTPTKMEALARKMGDPSPLAVRLSNPTL SSKEQIQRHLLOHYNFREOIEEPDQICLTSAEYQIHHQSVLLHGERITKVRDILL KSYREGAFSSWLLTYGNQTYNFIYELFTLLPELTKITEMEKMPQRAVYTLASRQ GPOKKEEIRNRYRGERKSELLDRKKEFFLVETDCRKTSPVQOALAMLTGKSQILTK CTSLSDQEITLLEKLIKLEKVKSNLFPDTKV" complement(9685..10452) /gene="mind" /complement(9685..10452) /gene="mind" /codon_start=1 /transl_table=11 /product="chromosome partitioning ATPase" /protein_id="BAA99013.1" /db_xref="GI:8979179" /translation="MKTIAVNSFKGSGTAKTSTTLHLGAALAOYHOARVLLIDFDAQN LTSGLGLDPCYDSLAVVLQGEKEIQEVI RPIQDTOLDLIPADTWLERIEVSGNLAAD RYSHERLKYVLGSDYQDYVILIDTPPSICHLTSAIADYALICATPEYFSVKGLE RYLAGFIQISARPLILQVALSVNCRKNSAFELIHKTFPGKLLNTKIRRDITY SEAAHGRPVFATSPSARASEDYFNLTRELLILLADI" complement(10485..12392) /gene="thrS" /complement(10485..12392) /gene="thrS" /codon_start=1 /transl_table=11 /product="threonyl tRNA synthetase" /protein_id="BAA99014.1" /db_xref="GI:8979180" /translation="MIQVTCQDNKYEVLEGTAAELAKOLKNSHQFTGVLINERPRDL STHLNEGDTLVLTSEDPEGREIFLHTSAHLLAQAVLRWLPDAIPTIGVYIDHGFYVD FANLISSEDFVLIEDTVKQIVDEKLAISRETYGDKQQAQFPQNPFTELIRELPE NEETSAYSQGEFQIDLCRPHLSTAHKAFKVLRTSAAYWRGDPSPRESLVRYGTSFP TSKELRAHLQIEBAKKRDRHVLGAKLDLFSQESSGPMFPFHRGPIKPMVDALRYWK QHTAACYKEFLPOLMQLWEVSHWQNDYKANNYTLQIDDEYAIKPMNCPGCMY YKTRLSHYKEFLURVAEVGHVRHQASGLMRVRAFHQDDAHVFLPPEOVEETL NIIQLNSTLYGTGFLHLELSTRPEKDTIGDLSLWELATDALNRALVOSGTFVVRP GEGAFYGPKIDIIHKDAIQRTWCCTIOLDMFLPERFELEYTTAOGTKSVPMVHLRAL	gene CDS	1889.00	Length: 362

Ratio: 5.218 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-438-185-1047 x AP002548/rev ..
Align seg 1/1 to reverse of: AP002548 from: 1 to: 325865
1 ValHisTyrCysGluArgThrLeuAspProLysTyrIleLeuLysIleAl 17
296378 GTGCACTACTGCGAGAGAAACCCCTGGACGCAAGTATATCTGAAGATTGC 296329
17 aLeuLysLeuArgGlnSerLeuSerLeuPhePheGlnAsnSerGlnSerL 34
296328 TCTAAAGCTGAGACAATCACTTTCCCTGCTTCTCCAGAACAGCAATCAC 296279
34 euGlnArgAlaTyrSerThrProTyrSerTyrTyrArgIleIleLeuGln 50
296278 TCCAACTGTCATCTCGACCCCATATCTCTACTACCGAATCAATCTACAA 296229
51 LysGlnAsnLysGlnLysGlnAlaLeuAlaArgHisLysCysIleSerL 67
296228 AAGGAAATAAAGAGAGCAAGCTTTAGCTGCACAAATGCATTTCTAT 296179
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
296178 TTTAGATATTTTCAAAACATCTCTTTGTTTCATCTTCTGTGCATATATCAA 296129
84 YAsnGlnArgGluGlyCysSerThrAspMetAlaValSerThrPro 100
296128 AGAATCAAAAGGAAGGTGCTCCACTGATATGCTGTGTGTAAGCACTCCC 296079
101 PhePheAsnArgAsnLeuTyrTyrArgLeuLeuSerSerArgPheSerLe 117
296078 TTTTAAATCGGAATTTATGCTATCGACATCTTCTCTCACCGTTTCTCT 296029
117 uTrpLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
296028 ATGGAAGAGCTATTGTTCCAAAGATTTTCTTCTGATTTAGTAAGCTTTCG 295979
134 lLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPhePheGlu 150
295978 GTCCTCTTCGATTTCTTAGACCATCAAGCAGTCATTAATTTCTCGAA 295929
151 LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProHi 167
295928 TTAGAACAACATTTTCTATATCCGTTTCAGATTTGAGTCTCCCA 295879
167 sGlnTyrLeuSerLeuGlnAspArgTyrPheProIleAlaSerValM 184
295878 TCAATACTGTCTCTGTGTCAGGACCGTTACTTCCCATTCCTCTGTAA 295829
184 eArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHis 200
295828 TGGGAACCTCGAATAAGATAATTTCTCTTAACTCTGATCTCATCCAT 295779
201 AspLeuGlyHisValProTyrPheLeuHisProSerPheSerGluPh 217
295778 GACCTTTTAGGCGACGTCCTTGGCTTCTACATCCCTCATTTTCTGAAT 295729
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGlnA 234
295728 TTTTCATAACATGGGAGACCTCTTCACTAAAGTCATAGAAAAAGTACAAG 295679
234 lAlaLeuProSerLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
295678 CTCTTCTAGTAAACACGACCATCAACACCTACAAAGCAATCTGATC 295629
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267
295628 GCATTTGAGCTGCTTTGTTGTTTACGTGTGTAAGCGGACTTATTGAAGA 295579
267 nHisGluGlyArgLysAlaTyrGlyAlaValLeuSerProGlnG 284

|||||
295578 CATGAAGGAAGAAAGCATATGAGCGTTCTTATCAGTTCTCCTCAGG 295529
284 lLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGluLeu 300
295528 AATTTGACACGCTTTTCATTGATAACGTACGTGTTCTCCTTTAGAAATTG 295479
301 AspGlnIleArgLeuProPheAsnThrSerThrProGlnGluThrIle 317
295478 GATCAGATATTCGCTTCTCCTTCAATACATCACTCAACACAGACACTT 295429
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334
295428 ATTTTCAATAAGACATTTTGTGTAACCTGGTAGAACCTCCTCAAAATTAG 295379
334 lUtrpMetLeuAspGlnGlyLeuGluSerIleProLeuTyrAsnGln 350
295378 AATGGATGCTCGACCAAGGCTCTGTAGAAATCAATTCCTCTTACAATCAA 295329
351 GluLysTyrLeuSerGlyPheGluValLeuCysGln 362
295328 GAGAAATATCTTCTGCTTTGAGGTACTTTGCCAA 295293
seq_name: gb_bal:AE004410
seq_documentation_block:
LOCUS AE004410 10143 bp DNA BCT 31-JUL-2000
DEFINITION Vibrio cholerae chromosome II, section 67 of 93 of the complete
chromosome
ACCESSION AE004410 AE003853
VERSION AE004410.1 GI:9658244
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 10143)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
REFERENCE 2 (bases 1 to 10143)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 10143
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="II"
/note="biotype: El Tor"
97. . 681
/gene="VCA0817"
97. . 681
/gene="VCA0817"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
gene
CDS

```

/gene="VCA0822"
complement(4464. .5888)
/db.xref="GI:9658245"
/translation="MSITVRYTMEYLLALLOTHYKWLMTILIVLYPHISRTILRLFES
MLTRGVDIRHKLAKRLIRTLIFLWFTTILWGLIELGLLWVGSLSFALIGVLGA
GWSLNVATLILFVQNCRIGCWVRIVDCAINFIEGKIIDMALLNVELETLEGNRLV
YPNNLFIIVRPVIVLQKAPEKAVKPERKQLPKGI"
684. .2048
/gene="VCA0818"
CDS
684. .2048
/gene="VCA0818"
/notes="similar to PID:1652477; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="magnesium transporter MgtE, putative"
/db.xref="GI:9658246"
/translation="MSEYQDLSQQOELIATAESEQTHLIGEMVENGDEGSIALILE
AFPEQVRMLRSLPLEMHIDVLTQMAEVRISVINALSEIKLTLAKLDNLSLIEW
EDSLPDIITSEALSINRDELELYDQAEADDEIGHWADRVKTYTLPLFSISVRRAKLV
LDKRYHDSQYIYVILINKAKKFGVLVSFSDVLAQETQLKRLMEELVTLNKQOTLSD
AVDALEHTLPIVPLVSDDETILGIDHWFALSQVREMYEARLMAGTGMDEGDLFAP
VIKSSKRGIMLIGINLLTALISITIGLFENVISQVVALVLMPIVASMGGIAGSOTL
TLVRAMALNOITVGNRFAILLKKNELGIGSINGLWALIGGMAALVFOEPVIGATIAL
AIVVNITTAALFGLVILIPILDKLPALAGSVILITGDCRGLFCLPRHQFGDVVA
AIRP"
2233. .2523
/gene="VCA0819"
CDS
2233. .2523
/gene="VCA0819"
/notes="similar to SP:P05380 GB:X07850 GB:X07899 PID:41613
PID:14616; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chaperonin, 10 kd subunit"
/db.xref="GI:9658247"
/translation="WNIRPLHDKLIVERLEVENKSEGGVILTSQSVKSNRGRKVVAVG
IGRPLKNGDRARMEVKTGDIIFNFGYKTEKVDGKEXYLILSESVDVAIVE"
2552. .4144
/gene="VCA0820"
CDS
2552. .4144
/gene="VCA0820"
/notes="similar to GB:U14003 SP:P06139 GB:X07850 GB:X07899
PID:146269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chaperonin, 60 kd subunit"
/db.xref="GI:9658248"
/translation="WAAKHVLFSTDAQKMLSGVNLNANAVKVLGPKGRHVVLNKSX
GAPTITKDSVAKETLEADKFEKMGQMLKQVAKNDEAGDGTATVLAQALINE
GMKAVANGNPNMOLKRGIDKAVSANVEKLHQLAKPCSDTQSTQVGAISANDHAIGE
YIQAQMEKVGKNGVITVEEQNALQNELSVYEGWQFDGHLSPFFINQPKAGCVLEENP
YVLVDKIKSHIELLPILLESVAKASRLLIITAEEDVDGALATFLVNSHRGIIKVAAY
KAFGEQKRMLEDIAALTAGRVISSEIGLELEKFTLDDLGSAKVTINKDNTITVD
KARPEALQDRTAQIQHQLHTTSOYDRDKLQQRIAKLSGGVAVIKIGATEVENKEK
KDRVDALHATRAAEVGEIVAGGVALIKIANELSLQGDNDQDNVGVIRALRAMEEP
LRQIINAGDEASVIANOVKTGDHYGYNAATGQFGNMLEMGILDPKAVTSLAQFOAA
SIAGLMTITTEAMVSEVDKESAA"
4177. .4383
/gene="VCA0821"
CDS
4177. .4383
/gene="VCA0821"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db.xref="GI:9658249"
/translation="WTKLPHQLSVVEAQPKTHPANAPMNSMSPQRCODRVLMPPYR
SVFPAVLVYVYINQAQQQPSPTLH"
complement(4464. .5888)

```

```

CDS
/gene="VCA0822"
complement(4464. .5888)
/notes="similar to GB:AE000782; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="aspartokinase, putative"
/db.xref="GI:9658250"
/translation="MMSYTVKEIGCTSMTADAVLENILLRPEOLYGRVFWVSAKSGM
TDALLECKRTGAGIYQIKAKHDESWRDVAYALEQRMILLINEMFADMTLRADAFYI
RARLEAIAICINILQICQIFQSLRQIPQIRREFLASGESHSAINALKNLKLGYN
ARFVDSLGCWCAENGLNDVIEEAFRDIDVTSELPIVGTASCOEGLMDRTDYGISEM
TFSRIAVTGAQQAIIHKETFLHSSADPRLVGAEKVRPIGTLTFEDVADQLANGLWEAIIH
PNAAGLRVQSNIDILRIKTFEPDHSGLTITASKEMNANSVTYYLSGSEPLNHRVQAERKYPQA
MVGMDKVSHELIEVTHAAKVLQIGKEMNANSVTYYLSGSEPLNHRVQAERKYPQA
ANGCVLHQALMPFAQPPSSARKVA"
complement(5885. .6301)
/gene="VCA0823"
CDS
complement(5885. .6301)
/gene="VCA0823"
/notes="similar to PID:2098610; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ectoine synthase"
/db.xref="GI:9658251"
/translation="MIYRLIECROSERRVVAENWESVRLMLKDDHMGFSFHTITIYA
NTQTHIHRNLESVYCMGEGEIEVVGKTYIPQPGTLYLDQHDHDLAFPSSEMV
MACVFNPLTGRHEIHDAEGVYPLDKSELISOCHREK"
complement(6366. .7601)
/gene="VCA0824"
CDS
complement(6366. .7601)
/gene="VCA0824"
/notes="similar to PID:2098609; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="diaminobutyrate--pyruvate aminotransferase"
/db.xref="GI:9658252"
/translation="MDIFKHESQVQSVYANHFPVLFGTAKGSWLYSQQDAYLDLFLSG
AGALNYGHNNAVLKQALLEYIERDGLTHGLDMHSEAKAHTQALQTHILEPRLGNYKL
OFTGPTGTNAVEAALKARKVTGRHNVYFTNGFHGCSGLAATGNHROGAGIAL
GVYRVPYDGYAGVDGLTLFETMLQDNSSGLDKPAAVLLETVOGEGNLVNSADWLQR
VQAICRAQOJILLIVDDIQAGCGRTGTFSPSGIIEPDMVTLKSLSGLGPLMALVLF
KPEWDMKPKGEHNGTFRGNHAFVATRALEYANODFOTHTAARSEQVTLAQCLL
SRVPTLPSGLKGLGMOGLACHNGDIARDIAALCFQKGLIETAGADEVLVKFCPLT
ITENDLAHGLTIIEVLLS"
complement(7625. .8146)
/gene="VCA0825"
CDS
complement(7625. .8146)
/gene="VCA0825"
/notes="similar to SP:052249; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="L-2,4-diaminobutyric acid acetyltransferase"
/db.xref="GI:9658253"
/translation="MIYPOIMHKPALPWVRRPRTQEDGLSIHIELIAQCAPLDQNSAYC
NFLQSSHFOYTCMLMAEQKLVGVFSAYRKPEOQNELFIQWAVHPSARGKGLAYQML
KHILAREDLADITVLETTITRSNQASWRLFOKLDREOGEQGSVSTFLDETCHFEHED
TEYLYRIPOQSSN"
complement(8259. .8375)
/gene="VCA0826"
CDS
complement(8259. .8375)
/gene="VCA0826"
/notes="identified by Glimmer2; putative"
/codon_start=1

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96724.1"
/db_xref="GI:9658254"
/translation="MLIRELAAYQPHRYFKIHFMFOFKVFLFLAKOFFSF"
complement(8846..9175)
/gene="VCA0827"

alignment_scores:
Quality: 256.50 Length: 254
Ratio: 1.666 Gaps: 6
Percent Similarity: 60.630 Percent Identity: 25.591
alignment_block:
US-09-438-185-1047 x AE004410/rev ..
Align seg 1/1 to reverse of: AE004410 from: 1 to: 10143
106 LeuTrpTyrArgLeuLeuSerSerArgPheSerLeuTrpLysSerTyCy 122
9890 GTTTGGCATGAGTATCACTCCGCGCAAGAGTGTGAAACCCAGAC 9841
122 sProArgPheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspP 139
9840 CTGCCAAGCTTATTAGATGGTTGAATATGCTTAATTTGCCGACAGATC 9791
139 heLeuAspHisGlnAlaValIle...LysPheGluLeuGluThrHis 154
9790 GCCTGCCACAATTCGCCGGAATAAAACCGAGTATTGCAAGAGAAACCGGT 9741
155 PheSerTyrTrpValSerGlyPheValAlaProHisGlnTyrLeuSe 171
9740 TGGCAAGTTGAGCTCTGCGCATTTGATGCTTTGATGCTTCTTTCG 9691
171 rLeuLeuGlnAspArgTyrPheProIleAlaSerValMetArgThrLeu 188
9690 CTTACTCGCCGATAAGAAGTTTCCAGTGCCACGCTTTTTCGCGGTAGAG 9641
188 sPLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeuGly 204
9640 AGAGTTTGATATCTCCAGAGCCGGATTTCTCCATGAAGTATGTT 9591
205 HisValProTrpLeuHisProSerPheSerGluPheIleAsnMe 221
9590 CACTCGCTATGCTCAGCCACCTGATTTGCGCTTTTACTCATGTTA 9541
221 tGlyArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerL 238
9540 TGGCCAGTAGTGTCTAAAGCCACCCCAAGAGCGTAGCTAC..... 9498
238 ysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIleAlaIleValArg 254
9497TTAGCCGCT 9489
255 CysPheTrpPheThrValGluSerGlyLeuIleGluAsnHisGluGlyAr 271
9488 CTGACTGTTTACGGTTGAGTTGCGTTAGTTTCAAGAACAGGACAAAC 9439
271 gLysAlaTyrGlyAlaValLeuIleSerSerProGlnGluLeuGlyHisA 288
9438 CAAATTTATGCGGGGCGCATTTTTCGTCACCGCGGTGAACCGTTTACG 9389
288 laphelleasAsnVal...ArgValLeuProLeuGluLeuAspGlnIle 303
9388 CTCGAGAAGACATATCCAAACGTCGACCGGTCGATATTCGACGATG 9339
304 lIeArgLeuProPheAsnThrSerThrProGlnGluThrLeuPheSerIl 320
9338 CTACGACCTCTTACCGATTGATATCATGACCGGATTTATTATGCTT 9289
320 eArgHisPheAspGluLeuValGluLeuThrSerLys.....L 333
9288 GCGGGATCTCAGTCAACTGATCACTCACTCACTCACTCACTCACTCACTGATGATGCGGT 9239

333 euGluTrpMet...LeuAspGlnGlyLeuLeuGluSerIleProLeuTyr 348
9238 TGGTGTGCAAGCTATGCAAGAGGGTACTCCG.....CCTCTTTT 9195
349 AsnGlnGluLys 352
9194 CAACCAAGGAA 9183
seq_name: gb_ov:GGA251387
seq_documentation_block:
LOCUS GGA251387 1633 bp mRNA VFT 02-DEC-1999
DEFINITION Gallus gallus mRNA for tyrosine hydroxylase (tyrosine hydroxylase gene).
ACCESSION AJ251387 GI:6523292
VERSION AJ251387.1
KEYWORDS tyrosine hydroxylase; tyrosine hydroxylase gene.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1633)
AUTHORS Ernsberger, U., Patzke, H., Tissier-Seta, J.P., Reh, T., Goridis, C. and Rohrer, H.
TITLE The expression of tyrosine hydroxylase and the transcription factors c-myc-2 and c-myc-1: evidence for distinct inductive steps in the differentiation of chick sympathetic precursor cells
JOURNAL Mech. Dev. 52 (1), 125-136 (1995)
MEDLINE 96076133
REFERENCE 2 (bases 1 to 1633)
AUTHORS Ernsberger, U.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1999) Ernsberger U., Ruprecht-Karls-University, Institut f. Anatomie und Zellbiologie III, Im Neuenheimer Feld 307, D-69120 Heidelberg, GERMANY
FEATURES
source
1. 1633
/organism="Gallus gallus"
/db_xref="taxon:9031"
/dev_stage="embryonic day 8"
/tissue_lib="sympathetic ganglia"
32. 1507
/gene="tyrosine hydroxylase"
32. 1507
/gene="tyrosine hydroxylase"
/codon_start=1
/product="tyrosine hydroxylase"
/protein_id="CA862388.1"
/db_xref="GI:6523292"
/translation="MPTPNISTSAKFRRAVSELDKQAEATMSPRFICRRQSLIED ARKREAAAAATAAESTETIVFEKDGRLNLFMLKGAKTSPLSRAKLVFEFEA KIHLETRLSKRPREGTALEVFVCEVHSDNLNTFTSSIKRVAEDVRTTKEDKHFWE PRKTCLELDKCHLVTKFDLPDLDPGYSDOVYRQRKSAIEAFHYKHGDDIPRVEY TAETATWKKVYSTLKSLYTHACKYLEAFNLEKFCGYNNENIPOLSEVSRFLKER TGFQLRVRAGLLSARDFLASLAFRVFOCTOYLIRHASSPMHSPEDCCHELLGHVPLA DKTAQFSODIGLASLCATDRELEKLTATYMTVEFGLCRQNGIVKAYGALLSSVGE LHSLSDEPEVRDFPDAAAVQPDONQVQPVFVSFSDAKNKLNRNYAAHKRPFS VKYEPYTHSLELDSPOTICHSLEVRDELHSLNALNVIS"
BASE COUNT 412 a 448 c 405 g 368 t
ORIGIN
alignment_scores:
Quality: 237.50 Length: 227
Ratio: 1.827 Gaps: 4
Percent Similarity: 57.269 Percent Identity: 29.075
alignment_block:
US-09-438-185-1047 x GGA251387 ..
Align seg 1/1 to: GGA251387 from: 1 to: 1633

[illegible]

```

183  TCATCTCCCAAGAAATGGC.....:
232  alGlnAlaLeuProSerLysGlnArgIleGlnThrLeuGlnSerAsn 248
1004  TGGCTCTCTGGAGCTTCAGAGGAGCGGTTCAGAACTGGAACG... 1050
249  LeuLeuAlaIleValArgCysPheTrpPheThrValGlnSerGlyLeuI 265
1051  .....TGCTACTCTTCTCACTGTGGAGTTTGGACTGTG 1082
265  eGluAsnHisGluGlyArgGlyAlaValGlyAlaValLeuLeuSerSerP 282
1083  CAGCAGATGGCAGCTTCAGAGCTTTGGTGGCGGCTTCCTTCTTCCA 1132
282  roGlnGluLeuGlyHisAla.PheIleAspAsnValArgValLeuProte 298
1133  TCAGTAGCTCAGACATGACCTTCCTGACATGCCAAGGTATAGCCCTTT 1182
298  uGluLeuAspGlnIleAlaArgLeuProPheAsnThrSerThrProGlnG 315
1183  G.....ATCCCAAGTTCCTGCAACACAGGAATGTCTCATCAC 1220
315  lu 315
1221  lu 1222

seq_name: gb_bai:AE004522

seq_documentation_block:
LOCUS AE004522 15724 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 83 of the complete
genome.
ACCESSION AE004522 AE004091
VERSION AE004522.1 GI:9946768
KEYWORDS Pseudomonas aeruginosa.
SOURCE Pseudomonas aeruginosa
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 15724)
Stover,C.K., Pham,X.-Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,F.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 15724)
Stover,C.K., Pham,X.-Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
source
location/Qualifiers
1..15724
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
complement(188..1387)
/gene="phhC"
/notes="phhC"
complement(188..1387)
/gene="phhC"
/codon_start=1

gene
complement(188..1387)
/gene="phhC"
CDS
complement(188..1387)
/gene="phhC"
/codon_start=1
```

```

/transl_table=11
/product="aromatic amino acid aminotransferase"
/protein_id="AAG04259.1"
/db_xref="GI:9946768"
/translation="MSHFQAVRPGDPTLGLLDVARNDRPADKLDLGVGVYKDAQGL
TPILRSVLAGORLVBQETPKSVYGGHGDALFAELALGAAKVSPLLEQRADATCT
PGTGLRLLAGDITLHCLPAGLWSDPTWPHETLEAAGLKVSHYPYVSADNRIDV
EMLAGLERIPQGVYLLHACCHNPFGFDLSHDDMFRLDVVRRELPLIDFAYQGF
GGLEEDAWAVRLFAGELPEVLYTSSCNFGLYRDRGVALIVCAONAKELITDLRSGL
AFIARNLMTSPHAGVAAIIGDSLKGMOEVEGMRSRISRLIGLVEALAPHL
LAERFAHVAQGRGMSFYTGSPQVAVRLRDEHAHYLVSSGRANVAGLDRRLDLAQA
IAQVCAD"
complement(1384..1740)
/gene="phhB"
/notes="PA0871"
complement(1384..1740)
/gene="phhB"
/codon_start=1
/transl_table=11
/product="pterin-4-alpha-carbinolamine dehydratase"
/protein_id="AAG04260.1"
/db_xref="GI:9946770"
/translation="NTALTOAHCEACRADAPHVSDLEPVLRLQIPDWNIEVRDGI
LEKYLKFKHALAFTNAVGEISEAGHHPLLTWGVKVTVTWWSHSIKGLHRNDFI
MAARTDEVAKTAEGRK"
complement(1845..2633)
/gene="phhA"
/notes="PA0872"
complement(1845..2633)
/gene="phhA"
/codon_start=1
/transl_table=11
/product="phenylalanine-4-hydroxylase"
/protein_id="AAG04261.1"
/db_xref="GI:9946771"
/translation="MKTQYVARQDDNGFIHPETEHQVWNLITRQLKLVIEGRACQ
FLDGIHQGLPHERIPQLEINRVLTQATGWVRVAFALIPQTFFELASQQPVA
TFITPEELDYLOEPIDEIHFHGLCLTNPFWAEFTHTYKGLKASKERVFARL
YMTIEFLVETDQGRYVGGISLPKRETYVLSDEPLHQAFLNPLEAMRTPTRIDIL
QPLYFVLRLDLKRLQLAQEDINALVHEAMRLGLHAPLFPKQAA"
2917..4476
/gene="phhR"
/notes="PA0873"
2917..4476
/gene="phhR"
/codon_start=1
/transl_table=11
/product="transcriptional regulator PhhR"
/protein_id="AAG04262.1"
/db_xref="GI:9946772"
/translation="MRKVKQNRVGIILRDILNLLVDYGINVRGEGVGGDGNATVLL
CPNMINLOSLRLKLENVGVYGVKRVGLMPSEHLELALLAALDPVLVSDVMGQ
GIVANRAAQLGLRVNDEVPGLPIRSRYEOLDLPELVANKARINGLRVKVGDFL
ADIALQSEHDESEALAGAVLILHRADNRVGRERIVYVKQELRGFDSIFOSSRMAAV
REARRMAPLADPLIIEETGTGKELLARACHLASPRGOSPMALNCAGLPESMAFEL
FYGFFPAGARPEGKGLLELTAGTFLPDGVGEMSPRLOAKLIPLODGGFRVGS
DEEYLDVRVTCATQVDLSLCAKGERODLTHRLNVLISLITPLIECLDGLAPLAEH
FLDQASQIGCGGPKLSAQALEKRYHWPGRVROLENYLFQAVSLCEGTYKAEHIR
LPDYGAQPLGDFSLGDDDAIVGRFEKAVLERFHEPSSRQLGKRRLGVSHHTAANK
LRQGVQSQSEG"
complement(4682..4948)
/gene="PA0874"
complement(4682..4948)
/gene="PA0874"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04263.1"
/db_xref="GI:9946773"
/translation="MKAFALMALFACSLPVMAGAQDQAPASQQENAVEEYAYGMOLDV
ARVISSTDTSNACGIVFAPMTYEDSNGKHTEYQVWNGNGCTG"
complement(5217..7412)
/gene="PA0875"
```


300 uAspGlnIleLeuArgLeuProPheAsnThrSerThrProGlnGluThrL 317
:::|::|::|::|::|::|::|
2012GGCATGCCGCAGCCCTACCGCATCGACATCTGTCAACCGCTCT 1970

317 eupheserlEargHisPheAspGluLeuValGluLeuThrSer..LysLe 333
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1969 ATTTCGTCTGCCGACCTCAAGCGCCTGTTCCACTGCCCCAGAAGAC 1920

333 uGluTrp 335
:::|::|
1919 ATCATGG 1913

seq_name: qb_pr10:HSWHYDR

seq documentation block: 1335 bp mRNA PRI 12-SEP-1993
 LOCUS HSMHYDR DEFINITION human mRNA for tryptophan hydroxylase (BC 1.14.16.4).
 ACCESSION X52836
 VERSION X52836.1 GI:37954
 KEYWORDS melatonin; serotonin; tryptophan 5-monooxygenase; tryptophan hydroxylase.
 SOURCE human.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1335)
TITLE	Boularand, S.
JOURNAL	Direct Submission
REFERENCE	Submitted (01-JUN-1990) Boularand S., CNRS, Bat. 32, Avenue de la Terrasse F91198 GIF sur Yvette, France
AUTHORS	2 (bases 1 to 1335)
TITLE	Boularand, S.; Darmon, M.C.; Ganem, Y.; Launay, J.M. and Mallet, J.
JOURNAL	Complete coding sequence of human tryptophan hydroxylase
MEDLINE	Nucleic Acids Res. 18 (14), 4257 (1990)
FEATURES	90332431 Location/Qualifiers

[illegible]

seq name: qb_ov:GGU26428

seq_documentation_block:			
LOCUS	GU26428	175 bp	mRNA
DEFINITION	Gallus gallus	tryptophan hydroxylase	mrna, complete cds.
ACCESSION	U26428		
VERSION	U26428.1	GI:1657695	
KEYWORDS			

```

MEDLINE 97072811
REFERENCE 2 (bases 1 to 1575)
AUTHORS Florez.J.C., Seidenman,K.J., Barrett,R.K., Sangoram,A.M. and
Takahashi,J.S.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1995) Neurobiology/Physiology, Northwestern
University, 2153 N. Campus Drive, Evanston, IL 60208, USA
FEATURES
source
1..1575
Location/Qualifiers
/organism="Gallus gallus"
/strain="White leghorn"
/sub_species="domesticus"
/db_xref="taxon:9031"
/tissue_type="pineal gland"
/clone="pTPH-37"
1..82
83..1420
/EC_number="1.14.16.4"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAC60036.1"
/db_xref="GI:3645921"
/translation="MIEDNKKNHAPERGRTAIFSLKNEVGLVKALKLQEKHVN
LVHIESKRNRSEFEIFVDCDSNRQLEIFQLKSHVSIYVMNPTHEFNVEDGD
MENIPWPKKISDLDCANRVLMDYGLDADHPGKDNVYKRRKYFADLANVYKHGD
PIPIETEEIKTWGTVRELKLYPHACREYLNPLTKYCGYRDNPQLEDV
SRFLKERTGTIRPVAGYLSPRDLAFLAVFHCTQYVRHSSDPLYTPDPOTCHELL
GHVPLAEPSTFAQSEIGLASLGADEAVOKLATCYFFTFEGLCKQSGOLRVYAG
LLSISLKHSLGSAKVPDPKVTCKQECGLITTFQEVYFVSESEAEKMRFAK
TIKRPFGKYNPTQSVQILDKSIASVNNELRHELDIVSDLSKMGKQLEV"
1421..>1575
3'UTR 433 a 368 c 381 g 393 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 231.00 Length: 229
Ratio: 1.819 Gaps: 4
Percent Similarity: 55.459 Percent Identity: 28.384

alignment_block:
US-09-438-185-1047 x GGU26428 ..
Align seg 1/1 to: GGU26428 from: 1 to: 1575
108 TyrArgLeuLeuSerSerArgPheSerLeuTriplysSerTyCysProAr 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
617 TACCGAGAGCTTAAC .....AAGCTTTACCCCACTCA 648
124 gPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeu 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
649 TGCGTCGAGAGAGTACTTAACACTTGCCTTCTCACCACCAATAGTGTG 698
141 sPHs .....GlnAlaValIleLysPhePhe 149
699 GGTCAGGAGAGCAACATATCCCGAGTCGAGAGAGTGTCCCGCTTCCTG 748
150 GluLeuGluThrHisPheSerTyTyTyProValSerGlyPheValAlaPr 166
||| : ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
749 AAAGACGCACAGGTTTACCACTGCGCAGTCCGCTGCTATCTGTCAAC 798
166 oHisGlnTyrLeuSerLeuLeuGlnAspArgTyTyTyPheProIleAlaSerV 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
799 CAGAGACTCTTGGCAGGATAGCATTACAGAGTTTTCACCTGCACTCAGT 848
183 alMetArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIle 199
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
849 ATGTTAGACACAGCTCGGACCTCTCTACACACAGAGAGCTGATACCTGC 898
200 HisAspLeuGluGlyHisValProTyrPheLeuHisProSerPheSerG1 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
899 CATGAGCTCTAGGCAATGTCCTCTTTTGCTGAACCCAGTTTGTGCTCA 948
```

```

216 uPhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValG 233
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
949 GTTCTCCAGGAAATGGT .....CTTG 971
233 InAlaLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeu 249
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
972 CATCATTGGGCGATCAGACGAGGCTGTCACAAAACCTGGCAACA..... 1015
250 IleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleG1 266
1016 .....TGCTACTTTTTCACGTAGAGTTTGGCCTGTGCAA 1050
266 uAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerProG 283
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1051 GCAAGGGGACAGCTTAGAGTTATGGGGCTGGCTCTCTCTCTATTATTA 1100
283 InGluLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGlu 299
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1101 GTGAGCTCAAGCACTCGCTCTCTGGCAGTGCCAAAGTCAAGCCTTTTGT 1150
300 LeuAspGlnIleLeuArgLeuProPheAsnThrSerThrProGlnGluTh 316
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1151 CCAAGGTCACCTGCAGCAAGAAAGTCCCTGTATACACTTTCCAGAGGT 1200
316 rLeuPheSerIleArgHisPheAspGluLeuValGlu 328
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1201 TTACTTTGTTCTGAAAGTTTGAAGAAGCAAGGAA 1237
seq_name: gb_rol:MUSTPHA
seq_documentation_block:
LOCUS MUSTPHA 1732 bp mRNA ROD 07-MAR-1995
DEFINITION Mouse tryptophan hydroxylase (Tph) mRNA, complete cds.
ACCESSION J04758.1 GI:202113
VERSION J04758.1 GI:202113
KEYWORDS tryptophan hydroxylase.
SOURCE Mouse (strain DBA; isolate P815) mastocytoma, cDNA to mRNA, clone
P813-113.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1732)
AUTHORS Stoll,J., Kozak,C.A. and Goldman,D.
TITLE Characterization and chromosomal mapping of a cDNA encoding
tryptophan hydroxylase from a mouse mastocytoma cell line
JOURNAL Genomics 7 (1), 88-96 (1990)
MEDLINE 90243261
COMMENT Draft entry and computer-readable sequence for [Genomics (1990) In
press] kindly submitted
by J.Stoll, 26-JAN-1990.
FEATURES
source
1..1732
Location/Qualifiers
/organism="Mus musculus"
/isolate="P815"
/strain="DBA"
/db_xref="taxon:10090"
/clone="pGT3-113"
/tissue_type="mastocytoma"
94..1717
/gene="Tph"
94..1437
/gene="Tph"
/EC_number="1.14.16.4"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAA63401.1"
/db_xref="GI:202114"
/translation="MIEDNKKNHKSHSRGRVTLIFSLNEVGLIKVLKIFQEN
HVSLLHIESKRKSRKORNSRSEFEIFVDCDSNRQLENDIFPLKSHATVLSVSDPDQLTAK
GDPIPKTEFTETEEIKTWGTVRELKLYPHACREYLNPLTKYCGYRDNPQLEDV
DVSNFLKERTGFSIRPVAGYLSPRDLAFLAVFHCTQYVRHSSDPLYTPDPOTCHE
LIGHVPLLAEPSTFAQSFQSEIGLASLGADEAVOKLATCYFFTFEGLCKQSGOLRVFG
```



```

389 GACACGGTTTCATCCACTATCCGGAACCGAGCACCAGGCTCGGAATAC 438
109 gLeuLeuSerSerArgPheSerLeuTyrPlySerTyrCysProArgPheP 126
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
439 CTTGATCCCGGCACTGAAGTGATCGAAGCGCGCTGTGAGGAAT 488
126 heLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis 142
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
489 ACCTCGACGGCATCGACAGCTCGC.....CTGCCCCAC 523
143 GlnAlaValIleLysPheGluLeuGlu.....Th 153
524 GACGGATCCCCAGCTCGAGATCAACAGGTTCTCCAGCGCCACCAC 573
153 rHisPheSerTyrTyrProValSerGlyPheValAlaProHisGlnTyrL 170
574 CGCTGGCGCGTGGCAGCGGTTCGGCGCTGATTCCTGTTCCAGACCTTCT 623
170 euSerLeuLeuGlnAspArgTyrPheProIleAlaSerValMetArgThr 186
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
624 TCGAAGTGTGGCCAGCAGCAATTCGCCGCTGCCACCTTCATCGGCACC 673
187 LeuAspLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeuLe 203
674 CGGAAGAACTGGACTACCTACCGAGCGCGGACATCTTCCACGAGATCTT 723
203 uGlyHisValProTyrPheLeuHisProSerPheSerGluPheIleA 220
724 CGGCCACTGCCACTGCTGACCAACCCCTGGCTCGCGGAGTTCCACCATA 773
220 snMetGlyArgLeuPheThrValIleGluLysValGlnAlaLeuPro 236
774 CTTACGGCAGCTCGGCTCAAGCG..... 799
237 SerLysGlnArgIleGlnThrLeuGlnSerAsnLeuIleAlaIleA 253
800 AGCAAGGAGGAACG.....GTGTCTCTCGC 825
253 lArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGluG 270
826 CGCCCTGTACTGATGACCATCGAGTTTCGGCCCTGGCGAGACCGCAGG 875
270 lArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnGluLeuGly 286
876 GCAAGCGCATCTACGGCGGCGCATCTCTCTCGCGGAGGAGACCGCTC 925
287 HisAlaPheIleAsp.....AsnValArgValLeuProLeuGluLe 300
926 TACAGCCTCTCGGACGAGCGCTGCACCGGCTTCAATCCGCTGGAG... 973
300 uAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrL 317
974 .....GGATGCGCGCGCCCTACCGCATCGACATCTTGCACACCGCTCT 1016
317 euPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLe 333
1017 ATTCTCTCTCGCCACCTCAAGCGCTTTCACACTGTCGCCAGGAGAAC 1066
333 uGluTrp 335
1067 ATCATGG 1073

```

seq_name: gb.pr10:HUMBTPH

```

seq_documentation_block: 1335 bp mRNA PRI 22-MAY-1995
LOCUS HUMBTPH
DEFINITION Homo sapiens tryptophan hydroxylase (Tph) mRNA, complete cds.
ACCESSION L29306
VERSION L29306.1 GI:531192
KEYWORDS tryptophan hydroxylase.
SOURCE Homo sapiens brain stem cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS Tipper,J.P., Citron,B.A., Ribeiro,P. and Kaufman,S.
TITLE Cloning and expression of rabbit and human brain tryptophan
JOURNAL Arch. Biochem. Biophys. 315 (2), 445-453 (1994)
MEDLINE 95077422
FEATURES
source
1..1335 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="E. coli DH5 (pJpHUMB)"
/tissue_type="brain stem"
1..1335
/gene="trph"
1..1335
/gene="trph"
/EC_number="1.14.16.4"
/notes="conflicts with Swissprot P17290 at amino acids 102,
151, 202, 207, and 390; protein phosphorylation at amino
acids 53, 58, 113, 260, 378, and 443; putative"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAA67050.1"
/db_xref="GI:531193"
/translation="MIEDNKENKDHSLRGRATLIFSLKNEVGLIKALKIFQEKHVN
LLHIESRKSRRNSEPIFVDCDTNREOLNDIFLLKSHNTNLSVTTPDNTMKKEGM
ESVWPFPKISLDLHCANRLMYGSELDAHPGKDNVYRKRKYADSNMSKYGDIP
IPKVEETEEIKTWGTVFRELNKLPHACREYLNKLPKLLSKYCGXQEDNIPQLEDIS
NFLKERTGSIPIVAGVLSRDFLSGLAFRVPHCTQYVHRSSDPFYTPEDTCHGLG
HYPLLAEPFAOFSOBIGLASGSEFAVOKLATCYFTVERGLCDQGLRVFGAGL
LSSISELKHVLSHARKVFPDKITCKQEGCLITFDQVYEVSESEFDEAKREFTKT
IKRFPGVKNPYSRIQLDKAKTITNANWELRDLDVSDALGKVSRRKPSI"
BASE COUNT 404 a 283 c 283 g 365 t
ORIGIN

alignment_scores:
Quality: 229.50 Length: 230
Ratio: 1.752 Gaps: 4
Percent similarity: 56.957 Percent identity: 27.391

alignment_block:
US-09-438-185-1047 x HUMBTPH
Align seg 1/1 to: HUMBTPH from: 1 to: 1335
107 TriTyArgLeuLeuSerSerArgPheSerLeuTyrPlySerTyrCysPr 123
||| : : : : : : : : : : : : : : : : : : : : : : : : :
520 TGGGAACCGTATTCGGGAGCTCAACAACTCTATCCGACCATGCTTG 569
123 oAtgPheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheL 140
||| : : : : : : : : : : : : : : : : : : : : : : : : :
570 CAGA.....GAGTATCTCAAAATTTTACCTCTGCTTCCCAAGTAT 610
140 euAspHisGlnAla.....ValIleLysPhe 148
611 GTGGATATCAGGAAGACAATATCCACAGCTGGAGATATTTCAAACTTT 560
149 PheGluLeuGluThrHisPheSerTyrTyrProValSerGlyPheValAl 165
661 TTTAAAGAGCGCACAGGTTTTTCCATCTGCTGTGGTGGTTACTTATC 710
165 aproHisGlnTyrLeuSerLeuGlnAspArgTyrPheProIleAlas 182
711 ACCAAGAGATTTCTTATAGGTTTAGCCCTTTTCGAGTTTTCACCTGC 760
182 erValMetArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeu 198
761 AATATGTGACACACAGTTCACACCCCTTCTATACCCACGACGCGGAT 810
199 IleHisAspLeuLeuGlyHisValProTrpLeuLeuHisProSerPheSe
||||| : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

seq_name: gb_ov:QULTYRHA
seq_documentation_block: 2077 bp DNA VRT 28-APR-1993
LOCUS QULTYRHA
DEFINITION Quail tyrosine hydroxylase mRNA, complete cds.
ACCESSION M24778
VERSION M24778.1 GI:213649
KEYWORDS tyrosine hydroxylase.
SOURCE Quail cDNA to mRNA.
ORGANISM Coturnix coturnix
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Coturnix.
REFERENCE 1 (bases 1 to 2077)
AUTHORS Fauquet M., Grima B., Lamouroux A. and Mallet J.
TITLE Cloning of quail tyrosine hydroxylase: Amino acid homology with
other hydroxylases discloses functional domains
J. Neurochem. 50, 142-148 (1988)
JOURNAL 88089590
FEATURES
     Location/Qualifiers
     1..2077
     /organism="Coturnix coturnix"
     /db_xref="taxon:9091"
     19..1494
     /note="tyrosine hydroxylase"
     /codon_start=1
     /protein_id="AAA49514.1"
     /db_xref="GI:213650"
     /translation="MPTNISAAKGFRAVSELDKQAEAIMSPFRIGRRQSLIED
     ARKEREAAAATDAESETTIVFEKQGRAMLNFPLMKGVKTSPLSRALKVFTFEA
     KIHLETRLSRKPREGTALEYFVRCVHSSDLNTFISIKRAVEDVTKEDKPHWF
     PKRICLDKCHLVTKFPDLDLHPGYSDQVYRQRKRSIAFAHYKHGDPPIRVEY
     TABETATWKEVSTLKSLYPTHACKYLEAFNLKFCGYENNPIPOLEEVSRLKVR
     TQGLRPVGLSLADRFVLAFOCTOYIRHASSPMHSPPCPCHELGHVPMLA
     DTKFAFSDIGLASIGATDEIEKLATWETVGLCRONGIVKAGAGLLSSYGE
     LIHLSDEPVEDPDPAACODPOYVVFVSEFSEDAKKNLRVNAHHRPFS
     VKYEPVTHSIELDSDPOTICHLSLESVDELHILNALNVIS"
BASE COUNT 553 a 556 c 485 g 483 t
ORIGIN
alignment_scores:
    Quality: 229.50      Length: 227
    Ratio: 1.807         Gaps: 4
    Percent Similarity: 55.947 Percent Identity: 29.075
ment_block:
us-09-438-185-1047 x QULTYRHA ..
Align seg 1/1 to: QULTYRHA from: 1 to: 2077
107 TTTTATATATCCGCCAGCATCTTCACCCATGCACCTCCCTGAGCGGACTGC 984
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
985 TGTCATGAGCTGCTGGCGCATGTCCTCAATGCTGCTGCTGACAGACATTCG 1034
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
215 rGluPhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysV 232
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1035 CCAGTTCTCCAGGACATGGG.....CTTGCACTTC 1066
: : : : : : : : : : : : : : : : : : : : : : : : : : :
232 alGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsn 248
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1067 TGGGAGCAACTGATGAAGAAATGAGAACTGGCAACACTT..... 1107
: : : : : : : : : : : : : : : : : : : : : : : : : : :
249 LeuIleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuI 265
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1108 .....TACTGGTTTACAGTGGAGTTGGACTCTG 1136
: : : : : : : : : : : : : : : : : : : : : : : : : : :
265 eGluAsnHisGluGlyArgLysAlaTyrglyAlaValLeuIleSerSerP 282
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1137 CACAGAATGGGATAGTCAAGCTATGCGGCGAGGCTCTCTCTTCAT 1186
: : : : : : : : : : : : : : : : : : : : : : : : : : :
282 roGlnGluLeuGlyHisAlaPheIleAsnValArgValLeuProLeu 298
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1187 ACGGGAGCTGATACACTCCTTGTGCATGAACACAGAGTACGGGACTTT 1236
: : : : : : : : : : : : : : : : : : : : : : : : : : :
299 GluLeuAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnI 315
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1237 GATCCTGAGCGTCTGCGGTTTCAGCCCTGCCAGGACGACGCTTACCAC 1286
: : : : : : : : : : : : : : : : : : : : : : : : : : :
315 uThrLeuSerIleArgHisPheAspGlu 325
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1287 TGTGTTTGTGTCTGAGAGCTTCAGTGAT 1317
: : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: gb_rol:MUSTHRA
seq_documentation_block: 1757 bp mRNA ROD 27-APR-1993
LOCUS MUSTHRA
DEFINITION Mouse tyrosine hydroxylase, complete cds.
ACCESSION M69200
VERSION M69200.1 GI:201997
KEYWORDS oxygen oxidoreductase; tyrosine 3-monooxygenase; tyrosine
hydroxylase.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1757)
AUTHORS Ichikawa, S., Sasaoka, T. and Nagatsu, T.
TITLE Primary structure of mouse tyrosine hydroxylase deduced from its
cDNA
JOURNAL Biochem. Biophys. Res. Commun. 176, 1610-1616 (1991)
MEDLINE 91248263
FEATURES
     Location/Qualifiers
     1..1757
     /organism="Mus musculus"
     /strain="ICR"
     /db_xref="taxon:10090"
     /dev_stage="7 wk old"
     /tissue_type="brain"
     <1..1757
     /gene="TH"
     1..1757
     /gene="TH"
     6..1502
     /gene="TH"
     /EC_number="1.14.16.2"
     /codon_start=1
     /product="tyrosine hydroxylase"
     /protein_id="AAA40434.1"
     /db_xref="GI:201998"
     /translation="MPTPSASSPQKGFRAVSEDTKQAEAVTSRPTFGRQSLIED
     ARKEREAAAAAAAAAVASAEPCNPLEAVVFEEDGNVALLFLSLRGTKPSSLSRLK

```

FETFEAKIHHELETRPQAQPLAGSHPLEYFVREVPESGLAAILSVRPVSDVRSAS
 KQELKQVWPKVFKVSDCHHLVTKFPDIDLDHLEHAFQRRKLNELNAFYKQGE
 LPIHPIVETKEETWVYATLKGIIYACHRHELSAFOYRRKLGREDSIQLEDY
 HIKHKTGRTQFQVAGLLASGLASDAETKVFCTQYIRHASSPMHSPEQDCHELL
 GHVMPDLRTPAFQSOQIGLASGLASDAETKLSVYFVFEFGKQNGELKAYGAG
 LLSVGLSSKLLHSSEEPVRAQVAFQVQVQYVQVYVYFVSFSDAKDKUNYAS
 RIOPRSPGKSDPPTYLAVDLDSPHTIRRSVQEQDELHITLQALSAE

BASE COUNT	368 a	531 c	472 g	386 t
ORIGIN				

alignment_scores:		
Quality:	225.50	Length: 240
Ratio:	1.658	Gaps: 5
Percent Similarity:	56.667	Percent Identity: 27.500

```
alignment_block:  
US-09-438-185-1047 x MUSTHRA ..  
  
.....in seq 1/1 to: MUSTHRA from: 1 to: 1757
```

[illegible]

Mon Jul 2 08:46:42 2001

us-09-438-185-1047.rge

Page 20

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:48:36 ; Search time 58.66 Seconds
(without alignments)
816.475 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHCERTLDPKYIKALKL.....ESIPLYNQKYLGSFEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_plant:**

10: sp_rodent:**

11: sp_unclassified:**

12: sp_vertebrate:**

13: sp_virus:**

14: sp_virus:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1886	99.8	362	2	Q9Z6L3	Q9Z6L3 chlamydia p
2	256.5	13.6	289	2	Q9KLB8	Q9KLB8 vibrio chol
3	237.5	12.6	491	13	Q9PU40	Q9PU40 gallus gall
4	224.5	11.9	497	5	Q96370	Q96370 schistosoma
5	219	11.6	579	5	Q24000	Q24000 drosophila
6	206.5	10.9	532	5	Q9XZD1	Q9XZD1 caenorhabdi
7	206.5	10.9	575	5	Q23438	Q23438 caenorhabdi
8	196.5	10.4	297	2	Q9XC88	Q9XC88 chromobacte
9	181.5	9.6	457	5	Q9XYQ5	Q9XYQ5 caenorhabdi
10	178.5	9.4	450	5	Q96947	Q96947 geodia cydo
11	178.5	9.4	555	5	Q9W0K2	Q9W0K2 drosophila
12	177	9.4	495	2	Q9W0K2	Q9W0K2 myxococcus
13	157	8.3	438	5	Q17498	Q17498 branchiosto
14	109.5	5.8	1253	10	Q9SCZ3	Q9SCZ3 arabidopsis
15	106.5	5.6	506	8	Q9GFC3	Q9GFC3 arabis alpi
16	106	5.6	129	13	Q42428	Q42428 lates calca
17	106	5.6	504	8	Q9GF55	Q9GF55 cardamine p
18	106	5.6	504	8	Q9GF31	Q9GF31 cardamine p
19	106	5.6	504	8	Q9GF30	Q9GF30 cardamine r

20	103.5	5.5	1782	5	Q9VPI9	Q9VPI9 drosophila
21	103	5.5	504	8	Q9GF41	Q9GF41 aubrieta de
22	103	5.5	513	8	Q9TIB8	Q9TIB8 bouteloua c
23	102.5	5.4	1154	5	Q9VMZ3	Q9VMZ3 drosophila
24	102	5.4	506	8	Q9GF64	Q9GF64 arabis alpi
25	101.5	5.4	519	8	Q9MW23	Q9MW23 oryza meyer
26	101	5.3	95	2	Q9RG34	Q9RG34 chromobacte
27	101	5.3	2042	5	Q9WIC5	Q9WIC5 drosophila
28	100.5	5.3	102	11	P97517	P97517 phodopus su
29	100.5	5.3	2212	14	Q9DQD1	Q9DQD1 ebola virus
30	99.5	5.3	513	8	Q9T197	Q9T197 sporobolus
31	99	5.2	924	11	O54921	O54921 rattus norv
32	98.5	5.2	513	8	Q9TIC0	Q9TIC0 zeugites pi
33	98.5	5.2	584	2	Q9PLZ8	Q9PLZ8 campylobact
34	98	5.2	1887	5	Q9ZG66	Q9ZG66 c contains
35	97.5	5.2	239	5	Q9NJQ3	Q9NJQ3 lymnaea sta
36	97.5	5.2	504	8	Q9GF36	Q9GF36 lepidium ca
37	97.5	5.2	519	8	Q9MW31	Q9MW31 oryza alta
38	97.5	5.2	519	8	Q9MW30	Q9MW30 oryza latif
39	97.5	5.2	519	8	Q9MW29	Q9MW29 oryza grand
40	97	5.1	347	4	Q13786	Q13786 homo sapien
41	97	5.1	504	8	Q9GF53	Q9GF53 arabis proc
42	97	5.1	504	8	Q9GF28	Q9GF28 arabidopsis
43	97	5.1	519	8	Q9MW25	Q9MW25 oryza ridle
44	97	5.1	519	8	Q9MW24	Q9MW24 oryza longi
45	96.5	5.1	363	8	Q9GHI8	Q9GHI8 caryodaphno

ALIGNMENTS

RESULT 1	
Q9Z6L3	
ID Q9Z6L3	PRELIMINARY; PRT; 362 AA.
AC Q9Z6L3;	
DT 01-MAY-1999 (TREMBLrel. 10, Created)	
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE AROMATIC AMINO ACID HYDROXYLASE.	
GN CPN1046 OR CPJ1045 OR CP0806	
OS Chlamydia pneumoniae (Chlamydia pneumoniae).	
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
OX NCBI_TaxID=83558;	
RN [1]	SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.	
RC STRAIN=CMW029;	
RX MEDLINE=99206606; PubMed=10192388;	
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;	
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";	
RL Nat. Genet. 21:385-389(1999).	
RN [2]	SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.	
RC STRAIN=J138;	
RX MEDLINE=20330349; PubMed=10871362;	
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,	
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;	
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138	
RT from Japan and CWL029 from USA.";	
RL Nucleic Acids Res. 28:2311-2314(2000).	
RN [3]	SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.	
RC STRAIN=AR39;	
RX MEDLINE=20150255; PubMed=10684935;	
RA White T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,	
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,	
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,	
RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,	
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;	
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	
RT pneumoniae AR39.";	
RL Nucleic Acids Res. 28:1397-1406(2000).	
RN [4]	EMBL; AE001685; AAD19183.1; -.


```
OY 159 PVSGFVAPHOYLSLLODRFYPIASVMKTLKDNFSLTPDLIHLLGHVPLWLLHPSFEFF 218
DB 283 PVAGLLSARDFLASLAFVFOCTOYIRHASSPMHSPPECCHELHGHVPLMLADTKTFAQS 342
OY 219 INNGRLFTVIEKVOALPSKKORTIOTLOSNIATVRCFWFTVESGLIENHGRKAYGAVL 278
DB 343 ODIG-----LASLGATEDEEKLATL-----YWFTEFGLCRQNGIVKAYGAGL 386
OY 279 ISSPQELGHAFIDNVRLPLELDQIIRLPNTSTPQETLFSIRHDE 325
DB 387 LSSYGLIHLSLDEPEVDFDPDAAVQYQDQYQPVYFVSEFS 433
```

```
RESULT 4
ID O96370 PRELIMINARY: PRT; 497 AA.
AC O96370;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
TRYPOTPHAN HYDROXYLASE (EC 1.14.16.4).
TPH.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Strigeidida; Schistosomatidae; Schistosomatidae;
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PUERTO RICAN;
RA Hamdan F.F., Ribeiro P.;
RT "Molecular cloning and characterization of tryptophan hydroxylase from
RL Schistosoma mansoni."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF031034; AAD01923.1; -.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; -.
DR Pfam; PF00351; bioperlin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOPHTERIN_HYDROXYL; 1.
KW Oxidoreductase.
SQ SEQUENCE 497 AA; 57598 MW; F8964E4B2C361D CRC64;
```

```
Query Match 11.98; Score 224.5; DB 5; Length 497;
Best Local Similarity 28.5%; Pred. NO. 9.7e-11;
Matches 68; Conservative 34; Mismatches 108; Indels 29; Gaps 5;
```

```
DB 96 VVSTPFNNMRYLLSRFLSKW-SYCPREFLDYLEAFGLLSDFLDH-----QAVI 146
DB 201 IVEYTEIKTGWIRYRELTLYKTSACHEF---QKNLGLLQKAGYNEFDLPQLQWS 256
OY 147 KFELETHESYTPVSGFVAPHOYLSLLODRFYPIASVMKTLKDNFSLTPDLIHLLGHV 206
DB 257 DFLKARTGCLRPVAGYLSARDLGLAFVFCYTOYIRHQPDPFTPEPCCHELLGHV 316
OY 207 PWLLHPSFEFFINMGRFLTIVKIEKVOALPSKKORTIOTLOSNIATVRCFWFTVESGLIE 266
DB 317 PMLADPKFARFSEIG-----LASLQTSDEEIKLAT-----CYFFTIEFGLCR 360
OY 267 NHEGRKAYGAVLISPPQELGHAFIDNVRLPLELDQIIRLPNTSTPQETLFSIRHDE 325
DB 361 ODNLKAYGAGLLSSVAELQHALSKAVIKFPMPKVINCECLVTFQNGYFEVSSFD 419
```

```
RESULT 5
ID Q24000 PRELIMINARY: PRT; 579 AA.
AC Q24000;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
```

```
DE PLE PROTEIN
GN PLE OR CG10118.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhakshi P., Bolshakov S.,
RA Borkov D., Botchan M.R., Bouch J., Brokstein P., Bottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C., Kodira A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liao X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Teh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA MEDLINE=95014502; PubMed=7929381;
RX Birman S., Morgan B., Anzivino M., Hirsh J.;
RT "A novel and major isoform of tyrosine hydroxylase in Drosophila is
RT generated by alternative RNA processing."
RL J. Biol. Chem. 269:26559-26567(1994).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA MEDLINE=90166583; PubMed=2483109;
RX Neckameyer W.S., Quinn W.G.;
RT "Isolation and characterization of the gene for Drosophila tyrosine
RT hydroxylase."
RL Neuron 2:1167-1175(1989).
DR EMBL; AE003561; AAF50648.1; -.
DR EMBL; U14395; AAG2876.1; -.
DR HSSP; P04177; ITOH.
DR FlyBase; FBgn0005626; ple.
DR InterPro; IPR001273; -.
DR Pfam; PF00351; bioperlin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOPHTERIN_HYDROXYL; 1.
```


Db 411 QFSQETG-----LASIGASEEDLKKLATL-----YFSTIERGLSDAADSPPVK 454
QY 266 ---ENHEGRKAYGAVLISSPOELGHAFIDNVRVLELDQIIRLPFNFTSTPQETLSIRH 322
Db 455 ENSGNHREKVVYAGLLSSAGELQHAVESATIIIRPDRVVEOCLITTFQSAFYFTNR 514
QY 323 FDELVELTSKLEWMLDQGLLESLPLYNOEKYLSGFEVL 360
Db 515 FEEAQO---KLRFMTNNMKRPFIYRN--PYTESVEVL 547

RESULT 8

Q9XC88 PRELIMINARY; PRT; 297 AA.
AC Q9XC88;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE PHENYLALANINE HYDROXYLASE.
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Chromobacterium.
NCBI_TaxID=536;
[1]
SEQUENCE FROM N.A.
RC STRAIN-ATCC 12540;
RA Volner A., Nersissian A.M., Abu-Omar M.M.;
RT "Expression, isolation, and metal-dependent catalysis of phenylalanine
hydroxylase from Chromobacterium violaceum."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF146711; AAD37774.1; --
DR HSSP: P00439; 4PAH.
DR InterPro: IPR001273; --
DR Pfam: PF00351; biopterin_H; 2.
DR PRINTS: PR00372; FWHYDRXLASE.
DR PROSITE: PS00367; BIOTERIN_HYDROXYL; 1.
SQ SEQUENCE 297 AA; 33594 MW; AAA67B609171FB0 CRC64;

Query Match 10.4%; Score 196.5; DB 2; Length 297;
Best Local Similarity 27.4%; Pred. No. 1.1e-08;
Matches 63; Conservative 40; Mismatches 94; Indels 33; Gaps 6;

QY 126 FLYLEAFGLSDFL-DHQAVIKFPELETHSYPSVSGFVAPHQYLSLLQDRYPIASVM 184
Db 63 FLEGLELEVDADRVDFNKLNEKMAATGWKIVAPGLIPDDVFFHELANRRFVTWML 122
QY 185 RTLDKDNFSLTPOLIHDLGHVPLLLHPSFEFFINMRLFTKVIEKVALPSKKRIQT 244
QY 123 REPHQLDYLOEPDVFHDLFGHYVPLLINPVFADYLEAYGKGVKR-KALGALP----- 173
QY 245 LQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQE---LGHAFIDNVRVLPLEL 300
Db 174 -----MLARLYNVTVEFGLINTPAGMRYGAGILSSKSEIYCLDSASPNRV---GFDL 224
QY 301 DQIIRLPNTSTPQETLSIRHFDVELTSKLEWMLDQGLLESIPLYNQ 350
Db 225 WEIMNTRYRIDTQKTVFVIDSKQLFDATAP-----DFAPLYLQ 264

RESULT 9

Q9AYQ5 PRELIMINARY; PRT; 457 AA.
AC Q9AYQ5;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PHENYLALANINE HYDROXYLASE (EC 1.14.16.1).
GN K08F8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;

RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RC "A phenylalanine hydroxylase gene from the nematode Caenorhabditis
elegans is expressed in the hypodermis."
RT J. Neurogenet. 0:0-0(1999).
KL EMBL: AF119386; AAD31643.1; --
DR HSSP: P00439; 4PAH.
DR InterPro: IPR001273; --
DR InterPro: IPR001273; --
DR Pfam: PF00351; biopterin_H; 1.
DR PRINTS: PR00372; FWHYDRXLASE.
DR PROSITE: PS00367; BIOTERIN_HYDROXYL; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 457 AA; 52189 MW; 7A573B884B9EF6FC CRC64;

Query Match 9.6%; Score 181.5; DB 5; Length 457;
Best Local Similarity 21.8%; Pred. No. 3.5e-07;
Matches 83; Conservative 66; Mismatches 137; Indels 95; Gaps 15;

QY 14 LKIALKROSLSLFFONSOSLORAY----STPVSYVRIILOKENKEKQALARKHCISILE 69
Db 81 LKTIIVKRELLSI---SNKLRKRFLEKGTGPKT-----KQKDSVPWFPOKINDIDQ 130
QY 70 FFKNLL-----FVHLLSKNQREGCSTDMA-----VVSTFFFNRL--WYRL 110
Db 131 FANRILSYGAELDADHPGKDMTYRERKFFADIAFNKHKIPIITITDEETATWTV 190
QY 111 LSSRFLW-KSYCPREFLDYLEAFGLLSD-----FLDHOAVIKFPELETHSYSPVS 161
Db 191 YNELTVMPKNACQEF--NYI--FPLQNCNGFGPDRIPQLODVSDFLKDCGTGTYIRVA 246
QY 162 GFVAPHQYLSLLQDRYPIASVMRTLDKDNFSLTPOLIHDLGHVPLLLHPSFEFFINM 221
Db 247 GLLSPRDLFAGWAFRVFHSIOYIRHSAKRYTPETPDICHELLGHVPLFADYVFAQSOEI 306
QY 222 GRLF----TKVIEKVALPSKKRIQTQLSNLIAIVRCFWFTVESGLIENHEGRKAYGAV 277
Db 307 GLASLAPDDVIEKLATL-----YWFTEFGICQDQGEKKAYGAG 346
QY 278 LISSPOELGHAFIDNVRVLPLE--LDQIIRLPNTSTPQETL----- 317
Db 347 LLSFGELOYALSDKPEVVDPAVCCVTYPIYEPYKPYFLAESFASAKNKLKSWAATI 406
QY 318 ---FSIRH--FDELVELTSKL 333
Db 407 NRFPQIRNAYTORVEILDV 427

RESULT 10

Q96947 PRELIMINARY; PRT; 450 AA.
AC Q96947;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PHENYLALANINE HYDROXYLASE.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Astrophorida; Geodiidae; Geodia.
OX NCBI_TaxID=6047;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99053257; PubMed-9877430;
RA Wiens M., Kozlowski C., Batel R., Mueller W.;
RT "Phenylalanine hydroxylase from the sponge Geodia cydonium:
implication for allorecognition and evolution of aromatic amino acid
hydroxylases."

```

RL Dev. Comp. Immunol. 22:469-478(1998).
DR EMBL: Y16353; CAA76184.1; -.
DR HSP: P00439; 4PAH.
DR InterPro: IPR001273; -.
DR InterPro: IPR001273; -.
DR Pfam: PF01842; bioterin_H; 1.
DR Pfam: PF01842; ACT; 1.
DR PRINTS: PR00372; FWHYDRLASE.
DR SEQUENCE 450 AA; 51204 MW; 691880218BB9725D CRC64;

Query Match          9.4%; Score 178.5; DB 5; Length 450;
Best Local Similarity 25.8%; Pred. No. 6e-07;
Matches 55; Conservative 43; Mismatches 80; Indels 35; Gaps 6;

QY 107 WYRLSSRFLSKSCPRFFLDYLAFLGLDLDLH-----QAVIKFFLETHFSY 158
D 186 WRTIFNLVDFPPTHACK---EHNHVPFLQENCGYREDNIPQLEEVSYQLSCGTGRLR 242
QY 159 PVSGFVAPHOYLSLDQRYFPFIASVMRTLDKNFSLTDPDLHDLGHVWPLLHPSSEFF 218
D 243 PVAGLLSRDFLAGLAFRVFHSYQYIRHSQNPYTPEDVCHLGHVCSV-ILFAQFS 301
QY 219 INMGLFTKVIKQVLPKSKQRIQIQLSNLAIIVRCFTWFTVSGLIENHGRKAYGAVL 278
D 302 QEIG-----LASGAPDEEYQQLAT-----LYWFTIERGLCKQDGQKAYGAGL 345
QY 279 ISSPQELGHAIDNVRVLPLEQLIIRLPFNTS 311
D 346 ISSFGELQYCLSDKPEVRPLD-----PFKTS 371

RESULT 11
Q9W0K2 ID Q9W0K2 PRELIMINARY; PRT; 555 AA.
AC Q9W0K2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG9122 PROTEIN.
GN CG9122.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kaloush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

```

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese H.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan K.C., Zhou X., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003470; AAF47444.1; -.
DR HSP: P04176; 1PHZ.
DR FlyBase: FBgn0035187; CG9122.
DR InterPro: IPR001273; -.
DR InterPro: IPR002912; -.
DR Pfam: PF00351; bioterin_H; 1.
DR Pfam: PF01842; ACT; 1.
DR PRINTS: PR00372; FWHYDRLASE.
DR PROSITE: PS00367; BIOTERIN_HYDROXYL; 1.
DR SEQUENCE 555 AA; 61530 MW; 60D527EFB1FA791C CRC64;

Query Match          9.4%; Score 178.5; DB 5; Length 555;
Best Local Similarity 32.0%; Pred. No. 7.9e-07;
Matches 47; Conservative 28; Mismatches 55; Indels 17; Gaps 3;

QY 143 QAVIKFFLETHFSYYPVSGFVAPHOYLSLDQRYFPFIASVMRTLDKNFSLTDPDLHDL 202
D 284 QDVSVLKRKTGTFQLRPVAGYLSRDLGLAFRVFHCYQYIRHSSDPFTPEPDCCHL 343
QY 203 LGHVPWLLHPSSEFFINMGRLFTKVIKQVLPKSKQRIQIQLSNLAIIVRCFTWFTV 262
D 344 LGHMPLLANSFSAFQSQEIG-----LASLGASDADIEKLATL-----YFFVVEF 387
QY 263 GLIENHGR-KAYGAVLISPOELGHA 288
D 388 GLCKQADSTFKYVAGLLSSVAELQHA 414

RESULT 12
Q9FDC3 ID Q9FDC3 PRELIMINARY; PRT; 495 AA.
AC Q9FDC3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE FAH.
GN FAH.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OC NCBI_TaxID=34;
[1]
SEQUENCE FROM N.A.
RA Sun H., Shi W.;
RT "Mrp, a genetic locus essential for cellular aggregation and
RT sporulation in Myxococcus xanthus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF285263; AAF99327.1; -.
DR SEQUENCE 495 AA; 53213 MW; CD82E52F8A613AEB CRC64;

Query Match          9.4%; Score 177; DB 2; Length 495;
Best Local Similarity 23.8%; Pred. No. 9.1e-07;
Matches 69; Conservative 47; Mismatches 128; Indels 46; Gaps 6;

QY 91 STDMAVSTPFPFNRLNRYLSSRFLSKSCPRFFLDYLAFLGLDLDLHQAIVKFFE 150

```

Db 31 SRDQV-----WRNLTSLRGLHADKAPVYLEGLEATGIGSECIPSLDENNEKL 80
 QY 151 LEHFSYVPVSGFVAPHOYLSLQDRFFPIASVMRTLDKDNFSLTPOLIHLGLHVPWLL 210
 Db 81 ARLGWACVGRGFIPAVFTELOQMGVLAADIRTHEIYETPAPDIVHESAGHAPIIA 140
 QY 211 HPSFSEFFINMGLRFTKIEKV--QAL-----PSKKQRIOTLQSNLIA-- 251
 Db 141 NRYAYELKACGLVGFKAISVQEQVFAIRNLVSVKDDPDASEEAAHAQARLEAASA 200
 QY 252 -----IVRCFWFVESGLTENHEGKAYGAVLISSPOELGHAFIDNVRVLPLELD 301
 Db 201 SRRVSESTRAGRLWTAETGLIGSVASPRYAGLFTSIGEAOHCLTPAVKKPLSV- 259
 QY 302 QIIRLPNTSTQETLFSIRHDELVELTSKLEWML-----DOGLESI 345
 Db 260 ACADMOYDITRMQPLFVARDFEHLFEVLAEFESTLSWKRGGDFGLTEAL 309

RESULT 13
 Q17498 PRELIMINARY; PRT; 438 AA.
 O17498;
 01-JAN-1998 (TEMBLrel. 05, Created)
 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE PHENYLALANINE HYDROXYLASE (EC 1.14.16.1).
 GN PAH.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7739;
 RN SEQUENCE FROM N.A.
 RA Patton, S.J., Luke, G.N., Holland, P.W.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DB EMBL; AJ001677; CRA04917.1; -;
 DR HSSP; P04176; iph7;
 DR InterPro; IPR001273; -;
 DR InterPro; IPR002912; -;
 DR Pfam; PF00351; Dioplerin_H; 2.
 DR Pfam; PF01842; ACT; 1.
 DR PROSITE; PS00046; HISTONE_H2A; UNKNOWN_1.
 KW Oxidoreductase.
 SQ SEQUENCE 438 AA; 49958 MW; 1619297DBDBF5EE7 CRC64;

Query Match 8.3%; Score 157; DB 5; Length 438;
 Best Local Similarity 24.1%; Pred. No. 3.7e-05;
 Matches 64; Conservative 37; Mismatches 129; Indels 36; Gaps 8;
 QY 105 NLWYRLSSRSLSMKSYCPR-----FFLDYLEAFGLSLDFLDH-QAVIKFFELTHFSYYP 159
 Db 170 NTRQFTLRKLYPTDACRHNYPVFLMNCGRFREDNIQLEDVSNFLKDCGTGFLRP 229
 QY 160 VSGFVAPHOYLSLQDRFFPIASVMRTLDKDNFSLTPOLIHLGLHVPWLLHPSFSEFI 219
 Db 230 VAGLLSSRDLAGLAFPPVHFSTQYIRHHSKPLYTPDPDVCHELLGHAPLADPSFAQFSQ 289
 QY 220 NGRLETKVIEKVALPSKKRIOTLQSNLIAIVRCFWFVTEGSLIENHIEGKAYGAVLI 279
 Db 290 EIG-----LASIGAPDDFVNMKIALT-----YWFTEFGLCRDQGEVKAQRDCC 333
 QY 280 SSPOELGHAFIDNVRVLPLELDQIIRLPNTSTQETLFSIRHDELVELTSKLEWMLDQ 339
 Db 334 HRSESCRYCLDKPDIRFEPEKISVTKYITEYQPIYFVADSFODAKEVR--TWS--- 388
 QY 340 GLLSEIPL-----YNOEKYLSGFEVL 360
 Db 389 ---HSIPRPFVSHYN--PYTOSVEIL 409

RESULT 14
 Q9SC23 PRELIMINARY; PRT; 1253 AA.
 AC Q9SC23; 2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-MAY-2001 (TEMBLrel. 16, Last annotation update)
 DE DISEASE RESISTANCE-LIKE PROTEIN.
 GN F26013.200.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN SEQUENCE FROM N.A.
 RA Delseny M., Berger C., Cooke R., Grillet F., Laudie M., Mewes H.W.,
 RL Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DB EMBL; AL133452; CAB63020.1; -;
 DR InterPro; IPR000157; -;
 DR InterPro; IPR000167; -;
 DR InterPro; IPR001611; -;
 DR InterPro; IPR002182; -;
 DR Pfam; PF00560; LRR; 5.
 DR Pfam; PF00931; NB-ARC; 1.
 DR Pfam; PF01582; TIR; 1.
 DR PRINTS; PR00364; DISEASERSIST.
 DR SMART; SM00255; TIR; 1.
 SQ SEQUENCE 1253 AA; 140809 MW; 4211ADE0566C2B01 CRC64;

Query Match 5.8%; Score 109.5; DB 10; Length 1253;
 Best Local Similarity 20.3%; Pred. No. 1.4; 130; Indels 127; Gaps 18;
 Matches 81; Conservative
 QY 77 VHLISLSKNQREGCS-----TDMVVSTPFFNNLWYRLISS-----RF 115
 Db 421 IHLI-----EGCGFFPRVEINVLVEKCLVSMA-EGRVVMHNLQSGIKLNGKRRS 472
 QY 116 SLMKSYCPRFFLDYLEAFG---LLSDFLDHOAV-----IKFEL-----ET 153
 Db 473 RLWKPLIKYFLEDQVLGSEDIEAIFLDPFSAISFDVNPMAFENMYNLYLKICSSNPGN 532
 QY 154 HFSYYPVSGFVAPHOYLSLQDRFFPIASVMRTLDKDNFSL-----TPD-- 197
 Db 533 HYALHLPKGVKSLPEELRLHWEHPFLSLPDQFNTNLVILNMCYSKLQRLWEGTKELG 592
 QY 198 -LIHLLGHVPWLLHPSFSEFFINM-----GRL-----FTKVI-----KVQ 233
 Db 593 MLKRIMLCHSQVLGVLQELQIALNMEVIDLQGCARLQFLATGHFOHLEVINLSGCIKIK 652
 QY 234 ALPSKKRIOTL---OSNLIAIVRCFWFVTEGSLIENHIEGKAYGAVLISSPOELG-HAF 289
 Db 653 SFPEVPPNIEELYKGTGIRSIPTVTFSPQDNSFYDHKDKHKLNFREVSSDSQSLSIMVY 712
 QY 290 IDNVRVLP-----LELQIIRLPNT-----STPQETLFSIRHDELVEL---TSKLEWM 336
 Db 713 LDNLKVLDSOCLELEDIQGIPNKLKLYLGTAIKELPSLMHLSLVLDLNCCKRLHK 772
 QY 337 LDOGL-----LESPLYNQEKYLSG 356
 Db 773 LPMGTGILSSLAIVNLNSGCGSELEDIQGIPRNLDEELIYLAG 811
 RESULT 15
 Q9GF63 PRELIMINARY; PRT; 506 AA.
 ID Q9GF63
 AC Q9GF63;

```

Query Match      5.6%; Score 106.5; DB 8; Length 506;
Best Local Similarity 22.6%; Pred. No. 0.75;
Matches 68; Conservative 37; Mismatches 131; Indels 65; Gaps 12;

QY 3 YCE-RTLDPKYILKIALKRLSCSLFQNSGSLQRAYSTPYSYRIILOKENKEKQALAR 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 YCNWKNFD---IKKLLILNPFFLYLFYHGVCE-----YESIFFLRKRSKHLRSTAY 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 HKCISILEFFKNL-----LFVH:-----LLISLKNQ-----RECGSTDMAVYSTP----- 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 EVLFERILFYAKIOHELKVFVNPPAILGLDKDPFLHYRVHGKSLATKDTPLLNKKWK 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 PFENLNLWRLSSRFSMLKWSYCPAFFLDYLEAGLLSDFLQHOQVAFKFEPELTHESYDPV 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 FYFNLWQFYFVSFQOSQIHNOLSKDNLEFLGYLSSLRNLPLVVSQMLENSFLI--- 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 161 SGFVAPHQYLSLQDQRYFFIASVMRTLDKDNFSLATPDLIHOLLG-----VPVLLHPFSFE 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 -----DNIRIKLONKPIPSISIGTSITKDKFC-----NLLGHPISKANW--TESSDSD 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 217 FFINMGRLFTKYIEKQVALPSKQR-----IQTLQSNLIAIYRCFVFTVEGSLI 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 ILNRFYRICNISHHYSQSSKKHLRYKIYILRLCCVTKLARKHKSTVRAFLKRLGSLGL 461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 266 E 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 E 462
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 29, 2001, 13:55:27
Job time: 411 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:40:51 ; Search time 40.35 Seconds
(without alignments)
683.400 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHCERTLDPKYLKIALK.....ESIPLYNQEKYLSGFEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1886	99.8	362	2 E72002	aromatic amino aci
2	1886	99.8	362	2 C86621	aromatic amino aci
3	256.5	13.6	289	2 D82413	phenylalanine-4-hy
4	235.5	12.5	262	2 F83535	phenylalanine-4-hy
5	231.5	12.3	444	1 S10489	tryptophan 5-monoo
6	229.5	12.1	262	2 A53452	phenylalanine hydr
7	229.5	12.1	444	2 S51159	tryptophan 5-monoo
8	229.5	12.1	491	2 A28582	tyrosine 3-monooxy
9	227.5	12.0	444	1 WHRTW	tryptophan 5-monoo
10	226.5	12.0	447	2 A34582	tryptophan 5-monoo
11	223.5	11.9	498	2 JX0088	tyrosine 3-monooxy
12	223.5	11.8	481	2 J15067	tryptophan 5-monoo
13	223.5	11.8	498	1 WHRT1	tyrosine 3-monooxy
14	221.5	11.7	491	2 I45983	tyrosine 3-monooxy
15	220	11.6	528	1 WHH04	tyrosine 3-monooxy
16	219	11.6	579	1 A55369	tyrosine 3-monooxy
17	214	11.3	453	1 WHRTF	phenylalanine 4-mo
18	213.5	11.3	452	1 S15758	phenylalanine 4-mo
19	212.5	11.2	493	2 J10039	tyrosine 3-monooxy
20	211.5	11.2	491	2 J00766	phenylalanine 4-mo
21	207	11.0	453	2 J00766	phenylalanine 4-mo
22	206.5	10.9	575	2 T34509	tryptophan 5-monoo
23	197	10.4	453	1 A42271	phenylalanine 4-mo
24	194	10.3	452	2 J04888	phenylalanine 4-mo
25	178.5	9.4	404	2 T25453	tyrosine 3-monooxy
26	175.5	9.3	457	2 T23494	phenylalanine 4-mo
27	109.5	5.8	1253	2 T45787	disease resistance
28	108.5	5.7	1501	1 B29813	174K nlnac protein
29	106.5	5.6	296	2 A40996	phenylalanine 4-mo

genome polyprotein
probable transcrip
exodeoxyribonuclea
NUP133 protein - y
exocyst complex pr
hypothetical prote
hypothetical prote
hypothetical prote
cytochrome P450 2L
exodeoxyribonuclea
glucose-6-P dehydro
glucose-6-P dehydro
probable membrane
exodeoxyribonuclea
large tegument pro
hypothetical prote

ALIGNMENTS

RESULT 1

E72002 aromatic amino acid hydroxylase - Chlamydomophila pneumoniae (strain CWL029)

C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C;Accession: E72002

R;Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A;Reference number: A72000; MUID:99206606

A;Accession: E72002

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-362 <ARN>

A;Cross-references: GB:AE001685; GB:AE001363; NID:q4377378; PIDN:AAD19183.1; PID:q437

A;Experimental source: strain CWL029

C;Genetics:

A;Gene: CPn1046

Query Match 99.8% Score 1886; DB 2; Length 362;
Best Local Similarity 99.7% Pred. No. 1.3e-144;
Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VHCERTLDPKYLKIALKQSLSEFQNSQSLQRAYSTPYSYRILLOKENEKQALA 60

Db 1 MHYCERTLDPKYLKIALKQSLSEFQNSQSLQRAYSTPYSYRILLOKENEKQALA 60

Qy 61 RHKCSILEFFKNLLFVHLLSLSKNQREGGCTDMVVSTPFFNRNLWYLLSSRFLSKWS 120

Db 61 RHKCSILEFFKNLLFVHLLSLSKNQREGGCTDMVVSTPFFNRNLWYLLSSRFLSKWS 120

Qy 121 YCPRFELDYLEAFGLSLDFLDHQAVIKFFETHTFSYYPVSGFVAPHQYLSLQDRYFPI 180

Db 121 YCPRFELDYLEAFGLSLDFLDHQAVIKFFETHTFSYYPVSGFVAPHQYLSLQDRYFPI 180

Qy 181 ASVMRTLDKDNFSLPDLTHLLGHVPWLLHPSFSEFFINMGRLFTKVIKQVQALPSKKQ 240

Db 181 ASVMRTLDKDNFSLPDLTHLLGHVPWLLHPSFSEFFINMGRLFTKVIKQVQALPSKKQ 240

Qy 241 RIOTLQSLNIAIVRCFWFTFVSGGLIENHEGRKAYGAVLISSSQELGHAFIDNVRVLPLEL 300

Db 241 RIOTLQSLNIAIVRCFWFTFVSGGLIENHEGRKAYGAVLISSSQELGHAFIDNVRVLPLEL 300

Qy 301 DOITRPFWTSTPQETLSIRHFDLVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360

Db 301 DOITRPFWTSTPQETLSIRHFDLVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360

Qy 361 CQ 362

Db 361 CQ 362

RESULT 2

C86621
aromatic amino acid hydroxylase [imported] - Chlamydomophila pneumoniae (strain J138)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: C86621
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: C86621
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-362 <STO>
A:Cross-references: GB:BA000008; NID:g8979419; PIDN:BAA99253.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CP11046

Query Match 99.8%; Score 1886; DB 2; Length 362;
Best Local Similarity 99.7%; Pred. No. 1.3e-144; Indels 0; Gaps 0;
Matches 361; Conservative 1; Mismatches 0;

QY 1 VHYCERTLPKYLKALKIROSLSLFFONOSLQRAYSTPSYIRILOKENEKQALA 60
DB 1 MHYCERTLPKYLKALKIROSLSLFFONOSLQRAYSTPSYIRILOKENEKQALA 60
QY 61 RHKCTISILEFFKNLLFVHLLSLSKNOREGCSMDMAVSTPFFNRLMYRLSSRFLSKS 120
DB 61 RHKCTISILEFFKNLLFVHLLSLSKNOREGCSMDMAVSTPFFNRLMYRLSSRFLSKS 120
QY 121 YCPREFDYLEAFGLSDFLDHOAVIKKPELETHFSYYPVSGFVAPHOYLSLQDRYFPI 180
DB 121 YCPREFDYLEAFGLSDFLDHOAVIKKPELETHFSYYPVSGFVAPHOYLSLQDRYFPI 180
QY 181 ASVMTLDKDNFSLTPDLIHDLGHVPMLLHPSSEFFINMGRFTKAVIEKVOALPSKKO 240
DB 181 ASVMTLDKDNFSLTPDLIHDLGHVPMLLHPSSEFFINMGRFTKAVIEKVOALPSKKO 240
QY 241 RIQTQSLNIAIVRCFWTFVESGLIENHGRKAYGAVLISSPQELHAFIDNVRVLPLEL 300
DB 241 RIQTQSLNIAIVRCFWTFVESGLIENHGRKAYGAVLISSPQELHAFIDNVRVLPLEL 300
QY 301 DQIIRLPNTSTPQETLSIRHFDDELVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360
DB 301 DQIIRLPNTSTPQETLSIRHFDDELVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360
QY 361 CQ 362
DB 361 CQ 362

RESULT 3
Phenylalanine-4-hydroxylase VCA0828 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82413
R:Heidelbergh, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Hardison, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: D82413
A:Status: preliminary
A:Molecule type: RNA
A:Residues: 1-289 <HEI>
A:Cross-references: GB:AE004410; GB:AE003853; NID:g9658244; PIDN:AAF96726.1; GSPDB:GN00142
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
C:Genetics:

A:Gene: VCA0828
A:Map position: 2

Query Match 13.5%; Score 256.5; DB 2; Length 289;
Best Local Similarity 25.6%; Pred. No. 2.4e-13;
Matches 65; Conservative 56; Mismatches 108; Indels 25; Gaps 6;
QY 106 LWYRLSSRFLSKSYCPREFDYLEAFGLSDFLDHOAVI-KFFELETHFSYYPVSGFV 164
DB 49 VVHELITRQEVVTRACQAYLDGLANLNLPTDRLPQLPEINRVLQRETGWQVPEPALI 108
QY 165 APOYLSLQDRYPIASVMRTLDKONFSLTPDLIHDLGHVPMLLHPSSEFFINMGR 224
DB 109 SDFREFFALLADKKFPVATFELRREEDYLPQEPDFHEVYGHGCMALTHPDFAAFTHYGQL 168
QY 225 FTKVIEKVOALPSKKQRIQTQSLNIAIVRCFWTFVESGLIENHGRKAYGAVLISSPQE 284
DB 169 GAKATPKERSY-----LRLYFVTEFGLVQEQOTKIYGGGILSSPGE 212
QY 285 LGHAFIDNV-RVLPLELDQIIRLPNTSTPQETLSIRHFDDELVELTSK----LEWM-LD 338
DB 213 TTYASESTIPREFDIMQVLRTPYRIDIMQPIYVLPDLSOLDSORDYMALVWQAMQ 272
QY 339 QGLLESIPLYNOEK 352
DB 273 DGLLP--PLFQPK 284

RESULT 4
F83535
Phenylalanine-4-hydroxylase PA0872 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83535
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
Adnan, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
Lozy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: F83535
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <STO>
A:Cross-references: GB:AE004522; GB:AE004091; NID:g9946768; PIDN:AA04261.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: phhA; PA0872

Query Match 12.5%; Score 235.5; DB 2; Length 262;
Best Local Similarity 24.5%; Pred. No. 1e-11;
Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;
QY 93 DMAVYSTPFFNRLMYRLSSRFLSKSYCPREFDYLEAFGLSDFLDHOAVIKKFELE 152
DB 13 DNGFIHYETHEQVWNTLIRQLKVIKGRACQYLDGIEQLG-----LPHRIQPLDEIN 67
QY 153 -----THFSYYPVSGFVAPHOYLSLQDRYPIASVMRTLDKONFSLTPDLIHDLGHV 206
DB 68 RVLTQATGWRVARVPALIPFQTFPELLASQCPVATFHTPEELDYLPQEPDFHEIFGHC 127
QY 207 PWLLHPSSEFFINMGRFTKAVIEKVOALPSKKQRIQTQSLNIAIVRCFWTFVESGLIE 266
DB 128 PLITNPWFAEFTHYTGKLGKA-----SKEER-----VFLARLYWMTTEFGIVE 171
QY 267 NHEGRKAYGAVLISSPQELHAFID--NVRVLPLELDQIIRLPNTSTPQETLSIRHF 323
DB 172 TQGGRIYGGGILSSPKETVYSLSDPELHQAFLNPLE---AMTTPYRIDILQPLVFLPDL 228
QY 324 DELVELTSK 332

Db 229 KRLFLAQAE 237

RESULT 5
S10489
tryptophan 5-monooxygenase (EC 1.14.16.4) - human
N:Alternate names: tryptophan 5-hydroxylase
C:Species: Homo sapiens (man)
C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 03-Mar-2000
C:Accession: S10489; S51559
R:Boularand, S.; Darmon, M.C.; Ganem, Y.; Launay, J.M.; Mallet, J.
Nucleic Acids Res. 18, 4257, 1990
A:Title: Complete coding sequence of human tryptophan hydroxylase.
A:Reference number: S10489; MUID:90332431
A:Accession: S10489
A:Molecule type: mRNA
A:Residues: 1-444 <BOU>
A:Cross-references: EMBL:X52836; NID:937954; PIDN:CAA37018.1; PID:937955
R:Tipper, J.P.; Citron, B.A.; Ribeiro, P.; Kaufman, S.
Arch. Biochem. Biophys. 315, 445-453, 1994
A:Title: Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA in
Reference number: S51199; MUID:95077422
A:Accession: S51559
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-18, 'T', '20-67, 'T', '69-89, 'TP', '92-96, 'M', '98-99, 'E', '101-103, 'S', '105-150, 'S', '15
35, 'G', '437-444 <TIP>
A:Cross-references: GB:L29306; NID:9531192; PIDN:AAA67050.1; PID:9531193
C:Genetics:
A:Gene: GDB:TPH: TPRH
A:Cross-references: GDB:120732; OMIM:191060
A:Map position: 11p15.1-11p14.3
C:Function:
A:Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetrahyd
A:Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; iron; melatonin biosynthesis; metalloprotein; monooxygenase; oxid
F:58/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status predict
F:260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #sta
F:272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.3%; Score 231.5; DB 1; Length 444;
Best Local Similarity 27.8%; Pred. No. 4.4e-11;
Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;
QY 107 WYRLSSRFSILWKSYPFRFLDYLEAFGLSLDFLDH-----QAVIKFELETHFSYY 158
Db 174 WGTVPQELNKLPHYTHACR---EYLNKPLLSKYCYREDNIPQEDISNFKLRTGFSIR 230
159 PVSQGVAPHOVLSLQDRYFPIASVMRTLDKNFSLTPDLTHDLGHVPLWLLHPSSEFF 218
Db 231 PVAGYLSPRDELGLAFRVFCHTCQVYVRHSSDPFYTPEDTCHELLGHVPLLAEPSPAQFS 290
QY 219 INMGRLFTKVIEKVOALPSKKRIQIQLQSNLIAIVRCFWFTVESGLIENHEGKAYGAVL 278
Db 291 QEIG-----LASIGASEAVQKLA-----CYFFTFVEGLCKDQGLRVFAGL 334
QY 279 ISGPQELGHADINRVLPLELQDIIRLPNTSTPQETLFSIRHDELVE 328
Db 335 LSSISLKHLSGHAHVPPDPKPKCKQECILITTFQDVYFVSESEFDEAKE 384

RESULT 6
A53452
phenylalanine hydroxylase - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 08-Oct-1999
C:Accession: A53452
R:Zhao, G.S.; Xia, T.; Song, J.; Roy, R.A.
Proc. Natl. Acad. Sci. U.S.A. 91, 1366-1370, 1994
A:Title: Pseudomonas aeruginosa possesses homologues of mammalian phenylalanine hydroxyl
A:Reference number: A53452; MUID:94151331

A:Accession: A53452
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-262 <RES>
A:Cross-references: GB:M88627; NID:9476740; PIDN:AAA25936.1; PID:9476741

Query Match 12.1%; Score 229.5; DB 2; Length 262;
Best Local Similarity 24.1%; Pred. No. 3.2e-11;
Matches 60; Conservative 52; Mismatches 104; Indels 33; Gaps 6;
QY 93 DMAVYSTFFNRLNWLSSRFSILWKSYPFRFLDYLEAFGLSLDFLDHQAIVKFELE 152
Db 13 DNGFIHYPETEHQVWNTLITROLKVIETRACQEQYLDGIEQLG-----LPHRIPOLDEN 67
QY 153 -----THFSYYPVSQGVAPHOVLSLQDRYFPIASVMRTLDKNFSLTPDLTHDLGHV 206
Db 68 RVLOATTGWRVARVPALIPQTFTELLASQQFPVATFRTPEELDYLOEPPDIFHEIFGHC 127
QY 207 PWLLHPSSEFFINMGRLFTKVIEKVOALPSKKRIQIQLQSNLIAIVRCFWFTVESGLIE 266
Db 128 PLITNPWLAEFTHTYVGLGLKA-----SKEER-----VFLARLYWMTIEFGLVE 171
QY 267 NHEGRKAYCAVLISSPQELGHAFID--NVRVLPLELQDIIRLPNTSTPQETLFSIRHF 323
Db 172 TDGKRIYGGGILSSPRETVYSLSDELPHQAENPLE---AMRTPYRIDILQPLYFVLPLD 228
QY 324 DELVELTSK 332
Db 229 KRLFLAQAE 237

RESULT 7
S51199
tryptophan 5-monooxygenase (EC 1.14.16.4) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 15-Jul-1995 #sequence_revision 19-Apr-1996 #text_change 31-Mar-2000
C:Accession: S51199; A32699
R:Tipper, J.P.; Citron, B.A.; Ribeiro, P.; Kaufman, S.
Arch. Biochem. Biophys. 315, 445-453, 1994
A:Title: Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA
A:Reference number: S51199; MUID:95077422
A:Accession: S51199
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-444 <TIP>
A:Cross-references: GB:L29305; NID:9531213; PIDN:AAA67051.1; PID:9531214
R:Greenett, H.E.; Ledley, F.D.; Reed, L.L.; Woo, S.L.C.
Proc. Natl. Acad. Sci. U.S.A. 84, 5530-5534, 1987
A:Title: Full-length cDNA for rabbit tryptophan hydroxylase: functional domains and e
A:Reference number: A32699; MUID:87289638
A:Accession: A32699
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-101, 'L', '103-150, 'L', '152-201, 'ND', '204-206, 'R', '208-389, 'K', '391-444 <GRE>
A:Cross-references: GB:M17250; NID:9165771; PIDN:AAA31487.1; PID:9165772
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; iron; metalloprotein; oxidoreductase; phosphoprotein
F:272,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.1%; Score 229.5; DB 2; Length 444;
Best Local Similarity 27.4%; Pred. No. 6.4e-11;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;
QY 107 WYRLSSRFSILWKSYPFRFLDYLEAFGLSLDFLDHQA-----VIKFELETHFSYY 158
Db 174 WGTVPQELNKLPHYTHACR---EYLNKPLLSKYCYREDNIPQEDISNFKLRTGFSIR 230
QY 159 PVSQGVAPHOVLSLQDRYFPIASVMRTLDKNFSLTPDLTHDLGHVPLWLLHPSSEFF 218
Db 231 PVAGYLSPRDELGLAFRVFCHTCQVYVRHSSDPFYTPEDTCHELLGHVPLLAEPSPAQFS 290

QY 219 INMGRLFTKVIKQALPSKKQRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 291 QEIG-----LASLGAEAEVQKLAT-----CYFFTFEGLCKQDQQLRVFGAGL 334
QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNSTPQETLFSIRHFDELVE 328
Db 335 LSSISLHSLHSGHAKVPDPKITYKQECLETTTQDVFVFSSEFEDAKE 384

RESULT 8
A28582
N: tyrosine 3-monooxygenase (EC 1.14.16.2) - quail
N: Alternate names: tyrosine 3-hydroxylase
C: Species: Phasianidae gen. sp. (quail)
C: Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 31-Mar-2000
C: Accession: A28582; PH1524
R: Pauquet, M.; Grima, B.; Lamouroux, A.; Mallet, J.
R: Neurochem. 50, 142-148, 1988
A: Title: Cloning of quail tyrosine hydroxylase: amino acid homology with other hydroxylase
A: Reference number: A28582; MUID: 88089590
A: Accession: A28582
A: Molecule type: mRNA
A: Residues: 1-491 <FAU>
A: Cross-references: GB:M24778; NID:G213649; PIDN:AAA49514.1; PID:G213650
R: Pauquet, M.; Boni, C.
J. Neurochem. 60, 274-281, 1993
A: Title: The quail tyrosine hydroxylase gene promoter contains an active cyclic AMP-resp
A: Reference number: PH1524; MUID: 93107923
A: Accession: PH1524
A: Molecule type: DNA
A: Residues: 1-30 <FA2>
C: Superfamily: phenylalanine 4-monooxygenase
C: Keywords: biotin; catecholamine biosynthesis; iron; metalloprotein; oxidoreductase;
F: 324, 329, 369/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.1%; Score 229.5; DB 2; Length 491;
Best Local Similarity 29.1%; Pred. No. 7.3e-11;
Matches 66; Conservative 34; Mismatches 100; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKYSCPRFFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
Db 226 MKVYSTLSKSLYPHACK---EYLEAFNLEKFCGYNNENNPQLEEFYRFLKERTGQFLR 282
QY 159 PVSGFVAPHQVLSLQDRYPPIASVMRTLDKDNFSLTDPDLHLLGHVPLLLHPSFEFF 218
1 283 PVYGLLSARDFLASLAFRVQCQYIRHASSPMHSPEDPCCHELLGHVPLMLADKTEAFQS 342
QY 219 INMGRLFTKVIKQALPSKKQRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 343 QDIG-----LASLGATDEIEKLATL-----YFTTFEGLCRQNGIVKAYGAGL 386
QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNSTPQETLFSIRHFDE 325
Db 387 LSSYGLHLSLSDPEVRDPPDAVAVQCQDPQYQVYFVSEFSFSD 433

RESULT 9
WHRTW
N: tyrosine 5-monooxygenase (EC 1.14.16.4) - rat
N: Alternate names: tryptophan 5-hydroxylase
C: Species: Rattus norvegicus (Norway rat)
C: Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 03-Mar-2000
C: Accession: J10034; A60034; A24367
R: Darmon, M.C.; Guilbert, B.; Levieil, V.; Ehret, M.; Mallet, J.
J. Neurochem. 51, 312-316, 1988
A: Title: Sequence of two mRNAs encoding active rat tryptophan hydroxylase.
A: Reference number: J10034; MUID: 88244702
A: Accession: J10034
A: Molecule type: mRNA
A: Residues: 1-444 <DAR>
A: Cross-references: GB:X53501; NID:957760; PIDN:CAA37579.1; PID:g57761
A: Experimental source: pineal gland

R: Kim, K.S.; Wessel, T.C.; Stone, D.M.; Carver, C.H.; Joh, T.H.; Park, D.H.
Brain Res. Mol. Brain Res. 9, 277-283, 1991
A: Title: Molecular cloning and characterization of cDNA encoding tryptophan hydroxylase
A: Reference number: A60034; MUID: 91245924
A: Accession: A60034
A: Status: not compared with conceptual translation
A: Molecule type: mRNA
A: Residues: 1-444 <KIM>
A: Experimental source: dorsal raphe nucleus
R: Darmon, M.C.; Grima, B.; Cash, C.D.; Maitre, M.; Mallet, J.
FEBS Lett. 206, 43-46, 1986
A: Title: Isolation of a rat pineal gland cDNA clone homologous to tyrosine and phenyl
A: Reference number: A24367; MUID: 87005247
A: Accession: A24367
A: Molecule type: mRNA
A: Residues: 167-261 <DAR>
A: Cross-references: GB:M28000; NID:G207432; PIDN:AAA42262.1; PID:G207433
C: Comment: This enzyme has different physical properties in pineal gland and in dorsa
C: Function:
A: Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetra
A: Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism
C: Superfamily: phenylalanine 4-monooxygenase
C: Keywords: biotin; iron; melatonin biosynthesis; metalloprotein; monooxygenase; o
F: 58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status pred
F: 260, 443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #
F: 272, 277, 317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 9.3e-11;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKYSCPRFFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
Db 174 WGTIFRELKLYPHACK---EYLRNPLLSKYCYGREDNVPOLEDVSNFLKERTGSIR 230
QY 159 PVSGFVAPHQVLSLQDRYPPIASVMRTLDKDNFSLTDPDLHLLGHVPLLLHPSFEFF 218
Db 231 PVAGYLSRDLGSLAFRVFCHTQYVRHSSDPLTYPTPEPTCHELLGHVPLLAEPFAQFS 290
QY 219 INMGRLFTKVIKQALPSKKQRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 291 QEIG-----LASLGAEAEVQKLAT-----CYFFTFEGLCKQDQQLRVFGAGL 334
QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNSTPQETLFSIRHFDELVE 328
Db 335 LSSISLHSLHSGHAKVPDPKVPACKQECLETSFDQVYFVSEFEDAKE 384

RESULT 10
A34582
N: tyrosine 5-monooxygenase (EC 1.14.16.4) - mouse
C: Species: Mus musculus (house mouse)
C: Date: 22-Jun-1990 #sequence_revision 09-Oct-1992 #text_change 31-Mar-2000
C: Accession: A34582
R: Stoll, J.; Kozak, C.A.; Goldman, D.
Genomics 7, 88-96, 1990
A: Title: Characterization and chromosomal mapping of a cDNA encoding tryptophan hydro
A: Reference number: A34582; MUID: 90243261
A: Accession: A34582
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-447 <STO>
A: Cross-references: GB:J04758; NID:G202113; PIDN:AAA63401.1; PID:G202114
A: Note: the authors translated the codon AAC for residue 405 as Gln
C: Superfamily: phenylalanine 4-monooxygenase
C: Keywords: biotin; iron; metalloprotein; oxidoreductase; phosphoprotein
F: 275, 280, 320/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 226.5; DB 2; Length 447;
Best Local Similarity 27.4%; Pred. No. 1.1e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

C:Species: Xenopus laevis (African clawed frog)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 31-Mar-2000
C:Accession: J51567
R:Green, C.B.; Besharse, J.C.
J. Neurochem. 62, 2420-2428, 1994
A>Title: Tryptophan hydroxylase expression is regulated by a circadian clock in Xenopus
A:Reference number: J51567; MUID:94246419
A:Accession: J51567
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-481 <GRE>
A:Cross-references: GB:120679; NID:9450644; PIDN:AAA21306.1; PID:9450645
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: biotin; iron; metalloprotein; oxidoreductase
F:309,314,354/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223.5; DB 2; Length 481;
Rest Local Similarity 29.5%; Pred. No. 2.2e-10;
Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;

QY 107 WYLLSSRFLSKVCPREFLDYLAFLGLSDFLDH-----QAVIKFELETHFSY 158
Db 177 WGTIFRELKLYPHACR---EYLNPLLSKVCYREDNIPQLEDVSNELKERTGFSIR 233
QY 159 PVSGFVAPHOYLSLQDRYPIASVMTLDKFNLSLTPDLTHDLGHVPMWLLHPSFSEF 218
Db 234 PVAGYLSRDFLUSGLAFVFCCTQYVHSDPLIPEPTDTCHELLGHVPLLAEPSPAQS 293
QY 219 INMGRLFTKVIKVOALPSKKQRIOTLOSMLIAIVRCFWFVSGSLIENHEGRKAYGAVL 278
Db 294 QEIG-----LASLGASEETVKLAT-----CYFTVEFGLCQDQGLRVFAGL 337
QY 279 ISSPQELGHAFIDNVRVLPFLDQIIRLPNTSTPQETLFSIRHFDLVE 328
Db 338 LSSISLKHLSLSEEPVAFDPFKIAKQCLITSFQDVYFVSESEDAKE 387

RESULT 11
JN0068
tyrosine 3-monooxygenase (EC 1.14.16.2) - mouse
A:Alternate names: tyrosine hydroxylase
C:Species: Mus musculus (house mouse)
C:Accession: JN0068; S21322
R:Chikawa, S.; Sasaoka, T.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 176, 1610-1616, 1991
A>Title: Primary structure of mouse tyrosine hydroxylase deduced from its cDNA.
A:Reference number: JN0068; MUID:91248263
A:Accession: JN0068
A:Molecule type: mRNA
A:Residues: 1-498 <ICH>
A:Cross-references: GB:M69200; NID:g201997; PIDN:AAA40434.1; PID:g201998
A:Experimental source: brain
R:Morgan, W.W.; Bermudez, J.; Sharp, Z.D.
submitted to the EMBL Data Library, June 1990
A:Description: DC-12 Nuclear Extracts Produce Tissue-Specific Protection of Several Sequences
A:Reference number: S21322
A:Accession: S21322
A:Molecule type: DNA
A:Residues: 1-30 <MOR>
A:Cross-references: EMBL:X53503; NID:g55055; PIDN:CAA37580.1; PID:g55056
C:Comment: This enzyme, which requires ferrous iron, catalyzes the hydroxylation of tyrosine in the physiology of adrenergic neurones.
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: biotin; iron; metalloprotein; monooxygenase; oxidoreductase; phosphoprotein
F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.9%; Score 225.5; DB 2; Length 498;
Rest Local Similarity 27.5%; Pred. No. 1.6e-10;
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;

QY 107 WYLLSSRFLSKVCPREFLDYLAFLGLSDFLDH-----QAVIKFELETHFSY 158
Db 233 WKEVYATLKLTAHACR---EHLERFQLLYCYGREDIPQLEDVSHFLKERTGQLR 289
QY 159 PVSGFVAPHOYLSLQDRYPIASVMTLDKFNLSLTPDLTHDLGHVPMWLLHPSFSEF 218
Db 290 PVAGLSARDFLASLAFRVFCCTQYIRHASSPMHSPEDPCHELLGHVPMWLLHPSFSEF 349
QY 219 INMGRLFTKVIKVOALPSKKQRIOTLOSMLIAIVRCFWFVSGSLIENHEGRKAYGAVL 278
Db 350 QDIG-----LASLGASEETVKLAT-----VYFTVEFGLCQDQGLRVFAGL 393
QY 279 ISSPQELGHAFIDNVRVLPFLDQIIRLPNTSTPQETLFSIRHFDLVE 328
Db 394 LSSYGLLSLSEEPVAFDPFKIAKQCLITSFQDVYFVSESEDAKE 453

RESULT 12
J51567
tryptophan 5-monooxygenase (EC 1.14.16.4) - African clawed frog

C:Species: Xenopus laevis (African clawed frog)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 31-Mar-2000
C:Accession: J51567
R:Green, C.B.; Besharse, J.C.
J. Neurochem. 62, 2420-2428, 1994
A>Title: Tryptophan hydroxylase expression is regulated by a circadian clock in Xenopus
A:Reference number: J51567; MUID:94246419
A:Accession: J51567
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-481 <GRE>
A:Cross-references: GB:120679; NID:9450644; PIDN:AAA21306.1; PID:9450645
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: biotin; iron; metalloprotein; oxidoreductase
F:309,314,354/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223.5; DB 2; Length 481;
Rest Local Similarity 29.5%; Pred. No. 2.2e-10;
Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;

QY 107 WYLLSSRFLSKVCPREFLDYLAFLGLSDFLDH-----QAVIKFELETHFSY 158
Db 211 WGTIFRELKLYPHACR---EYLNPLLSKVCYREDNIPQLEDVSNELKERTGFSIR 267
QY 159 PVSGFVAPHOYLSLQDRYPIASVMTLDKFNLSLTPDLTHDLGHVPMWLLHPSFSEF 217
Db 268 PVAGYLSRDFLUSGLAFVFCCTQYVHSDPLIPEPTDTCHELLGHVPLLAEPSPAQS 326
QY 218 INMGRLFTKVIKVOALPSKKQRIOTLOSMLIAIVRCFWFVSGSLIENHEGRKAYGAV 277
Db 327 SQEIG-----LASLGASEETVKLAT-----CYFTVEFGLCQDQGLRVFAGL 370
QY 278 LSSPQELGHAFIDNVRVLPFLDQIIRLPNTSTPQETLFSIRHFDLVE 334
Db 371 LSSISLKHLSLSEEPVAFDPFKIAKQCLITSFQDVYFVSESEDAKE 417

RESULT 13
WHRTY
tyrosine 3-monooxygenase (EC 1.14.16.2) - rat
A:Alternate names: tyrosine 3-hydroxylase
C:Species: Rattus norvegicus (Norway rat)
C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 03-Mar-2000
C:Accession: A00510; A44714; S03026; I58264
R:Grimaldi, B.; Lamouroux, A.; Blanot, F.; Faucon Biguet, N.; Mallet, J.
Proc. Natl. Acad. Sci. U.S.A. 82, 617-621, 1985
A>Title: Complete coding sequence of rat tyrosine hydroxylase mRNA.
A:Reference number: A00510; MUID:85113249
A:Accession: A00510
A:Molecule type: mRNA
A:Residues: 1-498 <GR>
A:Cross-references: GB:M10244; NID:g207408; PIDN:AAA42257.1; PID:g207409
R:Campbell, D.G.; Hardie, D.G.; Vulliamy, P.R.
J. Biol. Chem. 261, 10489-10492, 1986
A>Title: Identification of four phosphorylation sites in the N-terminal region of tyrosine hydroxylase.
A:Reference number: A44714; MUID:86278113
A:Accession: A44714
A>Status: preliminary
A:Molecule type: protein
A:Residues: 2-12;16-24;38-47;151-157 <CAM>
R:Bonnefoy, E.; Ferrara, P.; Rohrer, H.; Gros, F.; Thibault, J.
Eur. J. Biochem. 174, 685-690, 1988
A>Title: Role of the N-terminus of rat pheochromocytoma tyrosine hydroxylase in the regulation of its activity.
A:Reference number: S03026; MUID:88271342
A:Accession: S03026
A:Molecule type: protein
A:Residues: 2-26 <BON>
R:Harrington, C.A.; Lewis, E.J.; Krzemien, D.; Chikaraishi, D.M.
Nucleic Acids Res. 15, 2363-2384, 1987
A>Title: Identification and cell type specificity of the tyrosine hydroxylase gene promoter.
A:Reference number: I58264; MUID:87174758
A:Accession: I58264

A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-30 <RES>
A:Cross-references: EMBL:X04914; NID:g57355; PIDN:CAA28584.1; PID:g57356
C:Function:
A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylalanine
A:Pathway: catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: biotin; catecholamine biosynthesis; iron; metallopeptide; monooxygenase;
F:8/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status experiment
F:19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status exp
F:40/153/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status exp
F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223; DB 1; Length 498;
Best Local Similarity 26.7%; Pred. No. 2.5e-10;
Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSRFLSKS-----YCPRFLLYLEAFGLLSDFLDH-----QAVIKFELETHF 155
DB 227 AEIATWKEVYVTLKGLYATHACHREHLEGFQLLRYCGYREDSTIPQLEDVSRFLKERTGF 286

QY 156 SYTPSGVAPHOYLSLDRPPIASVMTLDKDNFSLTDPDLHLLGHVPMWLLHPSFS 215
DB 287 QURPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMWLLADRTFA 346

QY 216 EFFVNGRLFTKVLKQVLPSPKQRIOTLQSNLIAIVRCFTVSEFSFSAKDLKLSYASRIQ 446
DB 347 QFSODIG-----LASLGVSDEIEKLS-----VWFVTEFGLCKQNGELKAYG 390

QY 276 AVLLSSPOLGHAFIDNVRVLPLELDIIRLFNFTSTPQETLFSIRHF----DELVELTSKLE 334
DB 391 AGLSSYGEHLLSLEBEVRAFPDPTAAVQYQDQYQYVYFVSEFSFSAKDLKLSYASRIQ 446

QY 332 KLE 334
DB 451 RIQ 453

RESULT 14
I45983
tyrosine 3-monooxygenase (EC 1.14.16.2) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
A:Accession: I45983
R:Kobayashi, K.; Weisberg, E.P.; Stachowiak, M.K.; Turzai, L.M.; Gioio, A.E.; Kaplan, B.
J. Neurosci. Res. 19, 440-449, 1988
A:Title: Isolation and nucleotide sequence of a cDNA clone encoding bovine adrenal tyrosine
A:Reference number: I45983; MUID:88259287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: I45983
A:Molecule type: mRNA
A:Residues: 1-491 <DNX>
A:Cross-references: GB:X36794; NID:g163750; PIDN:AAA30779.1; PID:g163751
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: biotin; iron; metallopeptide; monooxygenase; oxidoreductase
F:324,329,369/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.7%; Score 221.5; DB 2; Length 491;
Best Local Similarity 27.1%; Pred. No. 3.2e-10;
Matches 65; Conservative 42; Mismatches 102; Indels 31; Gaps 5;

QY 107 WYRLSSRFLSKSVCPRFFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
DB 226 WKEVYVTLKGLYTHACK---EHLEAFELLERFCGYREDRIPQLEDVSRFLKERTGFQLR 282

QY 159 PVSGVAPHOYLSLDRPPIASVMTLDKDNFSLTDPDLHLLGHVPMWLLHPSFSSEFF 218
DB 283 PAAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMWLLADRTFAQFS 342

QY 219 INMGRLFTKVLKQVLPSPKQRIOTLQSNLIAIVRCFTVSEFSFSAKDLKLSYASRIQ 446
DB 343 QDIG-----LASLGVSDEIEKLS-----LYWTFVTEFGLCKQNGELKAYGAVL 278

QY 279 ISSPOLGHAFIDNVRVLPLELDIIRLFNFTSTPQETLFSIRHF----DELVELTSKLE 334
DB 387 LSSYGEHLLSLEBEVRAFPDPTAAVQYQDQYQYVYFVSEFSFSAKDLKLSYASRIQ 446

RESULT 15
WHUY4
tyrosine 3-monooxygenase (EC 1.14.16.2), splice form 4 - human
N:Alternate names: tyrosine 3-hydroxylase
N:Contains: tyrosine 3-monooxygenase, splice form 1; tyrosine 3-monooxygenase, splice
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jun-2000
A:Accession: A30002; A26825; A60201; JE0012; JE0013; JE0014; A27791; B27791; C27791;
R:Nagatsu, T. GenBank, December 1987
A:Reference number: A94509
A:Accession: A30002
A:Molecule type: mRNA
A:Residues: 1-528 <NAG1>
A:Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681
R:Kobayashi, K.; Kobayashi, K.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fuj
Biochem. Biophys. Res. Commun. 146, 971-975, 1987
A:Title: Isolation of a novel cDNA clone for human tyrosine hydroxylase: alternative
A:Reference number: A90136; MUID:87298614
A:Accession: A26825
A:Molecule type: mRNA
A:Residues: 1-94 <NAG2>
A:Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681
R:Le Bourdellès, B.; Boularand, S.; Boni, C.; Horellou, P.; Dumas, S.; Grima, B.; Mal
J. Neurochem. 50, 988-991, 1988
A:Title: Analysis of the 5' region of the human tyrosine hydroxylase gene: combinator
A:Reference number: A60201; MUID:88117543
A:Accession: A60201
A:Molecule type: mRNA
A:Residues: 1-65 <LER>
A:Cross-references: GB:M24790; NID:g556223; PIDN:AAA61174.1; PID:g556224
R:Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fuj
J. Biochem. 103, 907-912, 1988
A:Title: Structure of the human tyrosine hydroxylase gene: alternative splicing from
A:Reference number: JE0012; MUID:89008200
A:Accession: JE0012
A:Molecule type: DNA
A:Residues: 1-30, 62-135 <KOB1>
A:Cross-references: GB:D00269; NID:g220099; PIDN:BAA25094.1; PID:g2951764
A:Experimental source: splice form 1
A:Note: this splice form is produced by an alternative donor site within exon 1
A:Accession: JE0013
A:Molecule type: DNA
A:Residues: 1-34, 62-135 <KOB2>
A:Cross-references: GB:D00269; NID:g220099; PIDN:BAA25097.1; PID:g2951767
A:Experimental source: splice form 2
A:Accession: JE0014
A:Molecule type: DNA
A:Residues: 1-30, 35-135 <KOB3>
A:Cross-references: GB:D00269; NID:g220099; PIDN:BAA25095.1; PID:g2951765
A:Experimental source: splice form 3
A:Note: this splice form is produced by an alternative donor site within exon 1
R:Grima, B.; Lamouroux, A.; Boni, C.; Julien, J.F.; Javoy-Agid, F.; Mallet, J.
Nature 326, 707-711, 1987
A:Title: A single human gene encoding multiple tyrosine hydroxylases with different p
A:Reference number: A93393; MUID:87173064
A:Accession: A27791
A:Molecule type: mRNA
A:Residues: 1-30, 62-528 <GR11>
A:Cross-references: GB:X05290; NID:g32501; PIDN:CAA28908.1; PID:g32502
A:Experimental source: splice form 1
A:Note: this splice form is produced by an alternative donor site within exon 1
A:Accession: B27791
A:Molecule type: mRNA

A:Residues: 1-34,62-528 <GRI2>
A:Cross-references: GB:X05290; NID:g32501
A:Experimental source: splice form 2
A:Accession: C27791
A:Molecule type: mRNA
A:Residues: 30,35-528 <GRI3>
A:Cross-references: GB:X05290; NID:g32501
A:Experimental source: splice form 3
A:Note: this isoform is produced by use of an alternative donor site within exon 1
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: P05075; MUID:93371398
A:Accession: P05075
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 19-30 <ICH1>
A:Accession: P05082
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-61 <ICH2>
A:Accession: P05088
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 62-106 <ICH3>
R:O'Malley, K.L.; Anhalt, M.J.; Martin, B.M.; Kelsoe, J.R.; Winfield, S.L.; Ginns, E.I.
Biochemistry 26, 2910-2914, 1987
A:Title: Isolation and characterization of the human tyrosine hydroxylase gene: identification of a full-length cDNA clone encoding human tyrosine hydroxylase type
A:Reference number: I52396; MUID:88107612
A:Accession: I52396
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-61 <OMA>
A:Cross-references: GB:M18116; NID:g339633; PIDN:AAA77649.1; PID:g1004335
R:Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita
Nucleic Acids Res. 15, 6733, 1987
A:Title: Isolation of a full-length cDNA clone encoding human tyrosine hydroxylase type
A:Reference number: I38340; MUID:87316931
A:Accession: I38340
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-30,35-528 <KOB4>
A:Cross-references: EMBL:Y00414; NID:g37126; PIDN:CAA68472.1; PID:g37127
R:Ginns, E.I.; Rehavi, M.; Martin, B.M.; Weller, M.; O'Malley, K.L.; LaMarca, M.E.; McAL
J. Biol. Chem. 263, 7406-7410, 1988
A:Title: Expression of human tyrosine hydroxylase cDNA in invertebrate cells using a bac
A:Reference number: I55282; MUID:88213428
A:Accession: I55282
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-30,62-64 <GIN1>
A:Cross-references: GB:M20911; NID:g339636; PIDN:AAA611167.1; PID:g339637
A:Accession: I70056
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-34,62-64 <GIN2>
A:Cross-references: GB:M20912; NID:g339642; PIDN:AAA611168.1; PID:g339643
A:Comment: The expression of the four distinct proteins produced by alternate splicing v
C:Genetics:
A:Gene: GB:TH
A:Cross-references: GDB:119612; OMIM:191290
A:Map position: lplp5.5-llp15.5
A:Introns: 34/3; 61/3; 135/3
A:Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylalanin
A:Pathway: catecholamine biosynthesis
A:Note: this is the rate-limiting step in catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: alternative splicing; biotin; catecholamine biosynthesis; iron; metallopro
F:1-528/Product: tyrosine 3-monooxygenase, splice form 4 #status predicted <MAT4>
F:1-34,62-528/Product: tyrosine 3-monooxygenase, splice form 2 #status predicted <MAT2>
F:1-30,35-528/Product: tyrosine 3-monooxygenase, splice form 3 #status predicted <MAT3>

F:1-30,62-528/Product: tyrosine 3-monooxygenase, splice form 1 #status predicted <MAT
F:8/Binding site: phosphate (Thr) (covalent) (by unidentified kinase) #status predict
F:19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #statu
F:71,183/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status
F:361,366,406/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.6% Score 220; DB 1; Length 528;
Best Local Similarity 26.7%; Pred. No. 4.7e-10;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;
QY 112 SRSFSLWKS-----YCPREFLDYLEATGCLLSDFLDH-----QAVIKFFPLETHF 155
DB 257 AEETATWKEVYITLKGLYATHACGEHLEAFALLERTSGVREDNIPOLEDVSRFLKERTGF 316
QY 156 SYTPVSGFVAPHOYLSLLODRYPPIASVMRTLDKNFSLTPDLIHDLLGHVPMLLHPSFS 215
DB 317 QLRPVAGLLSARDEFLASLAFRVFOCTQYIRHASSPMHSPPEPCCHHELLGHVPMADRTFA 376
QY 216 EFFINMGRLETKVIEKVOALPSKKORIOTLOSNIATVRCFWFTVESGLIENHEGRKAYG 275
DB 377 QFSQDIG-----LASLGASDEEIEKLTSL-----WFTVEFLCKQNGEVKAYG 420
QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHF----DELVELTS 331
DB 421 AGLSSYVGLLHCLSEEPERAFADPEAAAVQPYODQTYQSVFVSEFSFSDAKDLRSYAS 480
QY 332 KLE 334
DB 481 RIQ 483

Search completed: June 29, 2001, 13:54:16
Job time: 805 sec
